To the Solar Foci

David Sonnabend

June 1, 1979

To the Solar Foci

David Sonnabend

June 1, 1979

ABSTRACT

Earlier authors (Cyranski and Lubkin, 1974) have shown that the sun is likely to act as a lens for gravitational radiation, with foci in the outer solar system. They have suggested that missions to these foci have the potential of directly measuring the density structure of the sun. Other applications include gravitational wave astronomy and new tests of general relativity.

The present work re-examines this idea, concentrating on the engineering aspects of focal missions - primarily spacecraft design and performance. Other topics studied include solar optics, gravitational wave detectors, navigation, and the design of missions for different purposes. Specifically, it will be shown that Shuttle launched chemical rockets have a substantial capability for reaching some foci; and that all can be reached with large payloads using nuclear isotope - electric propulsion. Thus, when gravitational wave detectors of sufficient sensitivity become available, a variety of new and attractive scientific missions will beckon.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Chapter

NOTATIONAL SYMBOLS

sx = sin x

cx = cos x

δx - variation in x

$\partial_x y = y_x = \partial y/\partial x$

$\hat{x}$ estimate of x

$\bar{x}$ indicates a 3-vector

x^T transpose of x

E(x) expectation of x

O(x) terms of order x

Einstein summation notation is used throughout in which Greek lower case indices have range 1, 2, 3; and Latin lower case indices have range 0, 1, 2, 3.

LATIN AND GREEK SYMBOLS

Symbol	Explanation	Equation of 1st Appearance
A	$c\theta_a$	III-9
	rib cross sectional area	III-93
	area swept out on an orbit	III-158
	area of radiator	V-1
	frontal area of corner cube	V-19
a	subscript-axial	
	transfer orbit semi-major axis	III-19
	constant in Van Der Waals equation	III-110
	edge of corner cube	V-18
	position error weight	VI-21
a_2, a_3	density cubic interpolation coefficients	A-2
B	$\partial_{yy}z$ (3 dimensional array)	IV-62
b	constant in Van Der Waals equation	III-110
	semi-minor axis	III-159
	control usage weight	VI-21
C	specific heat	V-2
	control gain matrix	VI-20
C_i	coefficients in power series	II-67
CL	compression failure limit	III-94
c	speed of light	V-5

Symbol	Explanation	Equation of 1st Appearance
D	subscript-departure	
	distance	III-83
D_c	corner cube effective diameter	V-20
DBL	diamond buckling limit	III-98
ds	line element	II-1
E	spacecraft specific energy	III-1
	Young's modules	III-95
	eccentric. anomaly	III-159
	heat energy	V-1
	electric field of laser	V-22
E_α^X	αth base vector in system X = S, G, N, T	III-4
E_0	amplitude of laser electric field	V-22
EL	Euler buckling limit	III-96
e	subscript-earth	
	transfer orbit eccentricity	III-16
	electronic charge	V-39
F	figure of merit	III-114
	beam return power fraction	V-37
	navigation plant matrix	VI-7
$F(r, \lambda)$	path equation function	II-22
$F(X, t)$	$\dot{X}$	IV-70

Symbol	Explanation	Equation of 1st Appearance
f	$[f_1, f_2]^T$ celestial mechanics constraints	III-132
	$\dot{x}$ (5 element vector)	IV-27
G	gain in focal region	II-105
	superscript - non-rotating earth coordinates	III-4
	corner cube gain	V-21
	control distribution matrix	VI-7
$G_j(R_0)$	function of R_0	II-65
g	acceleration of gravity	III-77
	$[g_1, g_2]^T$ departure and turn point constraints	III-136
	FM deviation of laser acoustic modulation	V-22
g_{ij}	metric tensor	II-2
H	hamiltonian	III-144
	observation matrix	VI-10
h	cubic deflection coefficient	II-74
	propellant level in tank	III-58
	integration step size	IV-71
	metric perturbation	V-5
$\overline{h}$	spacecraft specific orbital angular momentum	III-22

Symbol	Explanation	Equation of 1st Appearance
h_0	amplitude of metric perturbation	V-9
h_i	$x_{i+1} - x_i$	C-1
h_p	Planck's constant	V-39
I	transfer orbit inclination	III-14
	areal moment of inertia	III-95
	detector output current	V-39
I_c	nominal detector current for $\alpha = \pi/2$	V-40
I_{LN}	current due to laser noise	V-55
I_0	detector dark current	V-39
I_{SIG}	current due to gravitational wave	V-44
I_{SN}	current due to shot noise	V-43
I_{SP}	specific impulse	
I_{TN}	current due to thermal noise	V-52
J	performance index for ion trajectory	IV-19
	navigation performance index	VI-21
J^*	augmented performance index	IV-32
$J_i(g)$	Bessel function of the first kind	V-23
K	diamond buckling constant	III-97
	constant in time equation	III-160

Symbol	Explanation	Equation of 1st Appearance
K (cont.)	constant of axial motion	IV-11
	amplitude of noise component	V-30
	Kalman filter gains	VI-13
	time step constant	D-1
K_1, K_2, K_3	constants of integration	II-9-11
K_i	Runge-Kutta functions (i = 1 - 4)	IV-71
k	(n-1)/2	B-2
k_i	constants in payload function (i = 1 - 6)	III-140
L	rib length (axial or pyramidal)	III-81
	second stage payload	III-129
	detector baseline	V-11
	GC - F	VI-27
$L(0)$	$\ell(0)/r_s$	II-92
$L(R_0)$	$\ell(r_0)/r_s$	II-88
ℓ	transfer orbit semi-latus rectum	III-16
	length of cylindrical section of LH tank	III-69
	rib length (lateral)	III-80
	laser mirror spacing	V-25
$\hat{\ell}_a$	unit vector along axis	IV-18
$\ell(0)$	primary focal length of the sun	II-92

Symbol	Explanation	Equation of 1st Appearance
$\ell(r_0)$	focal length of sun for perihelion distance r_0	II-86
M	m/m_s	II-32
	mean anomaly	III-159
	$\partial_y z$ (matrix)	IV-51
	mass of detector	V-2
m	rib mass	III-99
	average spacecraft mass	IV-6
m_1, m_2	final stage mass	III-54
m_a	$m(t_a)$	IV-17
m_0	departure mass	III-54
m_p	propellant mass	III-54
m_R	mass of power supply	IV-1
$m_s = m(r_s)$	total mass-energy of sun	II-31
m_T	tank mass	III-65
m_{TC}	mass of cylindrical section of LH tank	III-73
$m(r)$	mass-energy of sun within radius r	II-26
Δm	extra mass for propulsion	IV-5

Symbol	Explanation	Equation of 1st Appearance
N	superscript of nodal coordinate system	III-5
n	number of moles	III-109
n_0	initial acceleration in g's	
P	$p/\bar{p}$	II-32
	compressive load in rib	III-88
	power available for propulsion	IV-2
	covariance of the estimator error	VI-14
P_A	additional power for propulsion	IV-5
P_B	power reserved for spacecraft systems	IV-2
P_D	total power at departure	IV-1
	power input to detector	V-38
P_E	power diverted to propulsion	IV-6
P_L	laser output power	V-37
P_x, P_y	beam return powers	V-37
p	pressure	III-64
p_h	hydrostatic pressure due to acceleration	III-60
$p(r)$	sun internal pressure	II-28
Q	$(1 - A^2 c^2{}_\phi)^{1/2}$	III-12

Symbol	Explanation	Equation of 1st Appearance
$Q(\alpha)$	shot noise function	V-47
q	normalized component of $\overline{v}_H$	III-38
	ion engine mass flow rate	IV-3
	$1/r$	IV-76
	process noise autocorrelation integral	VI-9
R	r/r_s	II-32
	universal gas constant	III-109
	photodiode operating resistance	V-51
R_0	r_0/r_s	II-38
$R^\Lambda(\theta)$	rotation about Λth axis through angle θ	III-4
r	Schwartzchild coordinate	II-1
	propellant tank inner radius	III-57
	radius at structural node	III-80
	tube radius of rib	III-93
	sensor noise autocorrelation integral	VI-11
$\overline{r}$	spacecraft heliocentric position	III-1
$\hat{r}_D$	heliocentric radial unit vector at departure	IV-20
r_e	mean distance of earth from sun	
$\overline{r}_e$	earth heliocentric position at departure	III-6

Symbol	Explanation	Equation of 1st Appearance
r_o	perihelion distance of ray	II-13
	Shuttle orbit radius	III-50
r_s	radius of sun	II-31
r_T	turnpoint distance	III-1
S	superscript for fundamental coordinate system	III-4
	control sweep matrix	VI-22
T	superscript for transfer coordinate system	III-15
	absolute temperature	III-109, V-1
	ion engine thrust	IV-3
T_{RL}	thrust of RL-10 engine	III-56
$T(R_0)$	time function	VI-4
t	Schwartzchild time coordinate	II-1
t_B	engine burn time	III-56
t_p	perihelion time	III-158
t_T	time at turnpoint	III-2
U	parabolic anomaly	III-166
$\hat{u}^a$	unit vector along axis	III-9

Symbol	Explanation	Equation of 1st Appearance
$\hat{U}^n$	transfer orbit normal unit vector	III-12
$\hat{U}^o$	Shuttle orbit normal unit vector	III-35
u	control vector	III-131
	control acceleration	VI-7
$\bar{u}$	ion engine acceleration	IV-14
V	tank volume	III-57
	corner cube volume	V-18
	photodiode reverse bias	V-51
v	specific volume	III-111
	position measurement noise	VI-10
$\bar{v}$	spacecraft velocity	III-1
$\bar{v}_a$	pre-turnpoint velocity	III-19
v_c	effective exhaust velocity	III-54
v_c'	adjusted effective exhaust velocity	III-127
$\bar{v}_D$	departure velocity	III-19
$\bar{v}_e$	earth velocity at departure	III-7
$\bar{v}_H$	departure earth relative hyperbolic excess velocity	III-37
v_h	normal component of $\bar{v}_D$	III-20

Symbol	Explanation	Equation of 1st Appearance
v_o	Shuttle orbit speed	III-50
v_r	radial component of $\bar{v}_D$	III-20
v_T	post-turnpoint speed	III-1
$v_{\shortparallel}$, $v_{\perp}$	axial, normal components of $\bar{v}_a$	III-21
Δv	velocity increment from electric propulsion	IV-6
Δv_D	departure impulse	III-50
Δv_T	turnpoint impulse	III-32
W	$\rho/\bar{\rho}$	II-32
$W_0(R)$, $W_2(R)$, $W_3(R)$	uniform, quadratic, cubic solar density models	II-78, 80
$W(\alpha)$	laser noise function	V-57
w	tank wall thickness	III-64
	rib wall thickness	III-93
	acceleration process noise	VI-7
	$(1-x^2)^{1/2}$	B-1
X	combined state of x and w	IV-69
	redefined 6-dimensional state vector	IV-72
	redefined 10-dimensional state vector	IV-76
	covariance of the state	VI-30
$\hat{X}$	covariance of the state estimate	VI-27

Symbol	Explanation	Equation of 1st Appearance
x	$[\sigma, \phi]^T$	III-131
	tan (E/2) or tanh (E/2)	III-162
	5-element state vector	IV-26
	navigation state vector	VI-7
x^i	generalized coordinates	II-2
Δx	distortion of baseline due to gravitational radiation	V-15
Y	normalized distance off axis (y/r_s)	II-94
	tensile or yield stress	III-64
y	$[\eta, \nu]^T$	III-131
	4 or 6 vector of initial guesses	IV-48
	distance off axis	A-8
y'	improved guess of y	IV-53
y_N	nominal value of y	IV-50
Δy	calculated error in y	IV-51
z	axial station in Shuttle bay	III-81
	subscript, cylindrical axial component	IV-23
	4 or 6 vector of terminal values	IV-49
	position measurement	VI-10
$\hat{z}$	cylindrical axial unit base vector	IV-23

Symbol	Explanation	Equation of 1st Appearance
z_D, Δz	desired value and error in z	IV-51
α	$3\gamma h/\delta_0'$	II-99
	r_e/r_T	III-26
	optical phase difference between baselines	V-35
	source right ascension	VI-1
α'	source ecliptic right ascension	VI-1
β	v_r/v_e	III-26
	electrical supply mass to power ratio	IV-1
Γ_{jk}^i	Christoffel symbol	II-3
γ	m_s/r_s	II-33
	v_h/v_e	III-26
	$(1 - \mu^2)^{1/2}$	V-40
Δ	angle of ray point from axis	II-93
δ	deflection	II-23
	increment in β and γ	III-155
	error in L from quarter wave	V-31
	laser noise fraction in signal band	V-55
	source declination	VI-1
δ'	source ecliptic declination	VI-1
δ_0, δ_2, δ_3	uniform, quadratic, cubic deflections	II-77, 79, 81
δ_e	external deflection	II-50

Symbol	Explanation	Equation of 1st Appearance
ϵ	alternative measure of deflection	II-16
	obliquity of ecliptic	III-4
	ullage fraction	III-59
	emissivity of radiator	V-1
	spatial strain	V-6
	laser strain	V-26
	frequency separation of noise from signal	V-30
	I_o/I_c	V-44
η	$\Delta v_D/v_e$	III-51
	ion engines electrical efficiency	IV-4
	detector quantum efficiency	V-39
θ	Schwartzchild coordinate	II-1
	$\frac{\pi}{2} - \lambda$	II-52
	transfer orbit true anomaly	III-16
	h/r	III-57
θ_a	focus ecliptic latitude	III-9
θ_D	true anomaly at departure	III-17
θ_K	Kennedy Spaceflight Center latitude	III-35
θ_T	true anomaly at turnpoint	III-17
$\Lambda(r)$	line element function	II-1

Symbol	Explanation	Equation of 1st Appearance
λ	angular representation of r	II-20
	$[\lambda_1, \lambda_2]$ Lagrange multipliers	III-141
	Lagrange multiplier or adjoint vector	IV-32
	wavelength of gravitational radiation	V-16
λ_1	$\lambda(1)$ or $\lambda(r_s)$	II-50
$\bar{\lambda}_{235}$	3 components of λ	IV-38
λ_L	laser wavelength	V-21
μ	gravitational constant of the sun	III-1
	beam return power asymmetry	V-37
μ_e	earth gravitational constant	III-50
ν	$\Delta v_T / v_e$	III-33
ρ	v_H / v_e	III-41
	propellant density	III-60
	subscript, cylindrical radial coordinate	IV-23
$\bar{\rho}$	average density of sun	II-31
	average second stage propellant density	III-106
$\hat{\rho}$	cylindrical radial base vector	IV-23
$\rho(r)$	mass-energy density of sun	II-27
σ	angular representation of r	II-15
	$v_{\shortparallel} / v_e$	III-26

Symbol	Explanation	Equation of 1st Appearance
σ (cont.)	Stefan-Boltzmann constant	V-1
σ_v, σ_w	sensor and process noise standard deviations	VI-12
τ	v_T/v_e	III-33
	power decay time constant	IV-2
τ_1, τ_2	constants in time delay relations	VI-2, 5
τ_v, τ_w	sensor and process noise correlation times	VI-12
$\Delta\tau(R_0)$	relativistic time delay in solar field	VI-2
$\delta\tau$	time difference between rays	VI-4
$\Phi(r)$	line element function	II-1
Φ_0	$\Phi(r_0)$	II-13
ϕ	Schwartzchild coordinate	II-1
	$\phi_a - \phi_e$	III-10
	ℓ/r	III-69
	subscript, cylindrical azimuthal coordinate	IV-23
$\hat{\phi}$	cylindrical azimuthal base vector	IV-23
ϕ_0	angle associated with q and $\overline{v}_H$	III-38
ϕ_1	$\phi_0 - \phi_e$	III-39

Symbol	Explanation	Equation of 1st Appearance
ϕ_2	critical longitude	III-49
ϕ_a	focus ecliptic longitude	III-9
$\hat{\phi}_D$	orbital velocity unit vector at departure	IV-20
ϕ_e	departure heliocentric longitude	III-5
$\Delta\phi$	optical phase shift due to gravitational wave	V-34
χ	v_o/v_e	III-51
ψ	trajectory parameter	II-3
	36 degrees (pentagonal angle)	III-80
	$\eta + \chi$	III-136
Ω	Shuttle orbit nodal longitude relative to E_1^G	III-35
ω	$[\omega_1, \omega_2]$ Lagrange multipliers	III-141
	rescaled λ	IV-40
	frequency of gravitational wave	V-9
$\bar{\omega}_{235}$	3 components of ω	IV-42
ω_m	acoustic modulation frequency of laser	V-22

A NOTE ON UNITS

The primary units adopted in this thesis are SI. This means that all distances will be in meters or their standard SI multiples. To get a feeling for these, some examples may help:

1 Mm = 621.37 miles ∿ distance from Stanford to Phoenix, Ariz.

1 Gm = 1.43682 solar radii = 2.6014 x mean lunar distance

1 Tm = 6.68459 astronomical units

Planetary mean heliocentric distances[1]:

Mercury	57.9 Gm	Saturn	1.427 Tm
Venus	108.2 Gm	Uranus	2.870 Tm
Earth	149.6 Gm	Neptune	4.497 Tm
Mars	227.9 Gm	Pluto	5.90 Tm
Jupiter	778.3 Gm		

The main variation from the SI is the use of gravitational units, in which $G = c = 1$, for the deflection analysis. With these, communication with professional relativists is noticeably easier. Other units are occasionally used to make contact with the literature.

INTRODUCTION AND SUMMARY

1. History

Regarding the sun as a gravitational lens is not new. That light rays passing close to the sun are bent is well established experimentally. The first correct prediction of this bending was by Einstein in 1916 (Ref. 37). The Einstein result is rederived here in greater detail in Section II-5. Earlier papers by Einstein, using only special relativity, predicted only half the bending. This same half-result can be found from Newtonian physics alone, and had been reported by earlier workers. Probably the first was a paper by J. G. Soldner in 1801. For a discussion of Soldner's work, see Ref. 51.

The Einstein result was that light from a distant star, passing by the edge of the sun would be bent inward by 1.75 arc seconds. It is easy to see that all such rays would be focussed to a point 82.01 Tm (or 548.2 astronomical units, or 13.9 times Pluto's distance) from the sun. To the author's knowledge, this focu has no practical value. Einstein, in later papers, considered the bending of one star's light by another, and predicted an unusual halo effect. However, the chance of finding such an alignment appears to be remote. Since then, others have looked at star clusters, galaxies, and the whole universe as gravitational lenses.

An interesting new idea was put forward in 1974 by Cyranski and Lubkin (Ref.23), who pointed out that the sun should be transparent to gravitational radiation. They then calculated the focal length of the center of the solar lens, reporting values of 3.41 and 4.42 Tm for two previously published solar density models. This work is confirmed here for these and three other solar models, and may be found in Appendix A. A difficulty with their paper is that their results

depend on a bending formula whose derivation is referred to only as "a lengthy and tricky analysis in curvature coordinates." The present Chapter II presumably parallels their analysis; while offering some extensions to their results.

Two other important points were made by Cyranski and Lubkin. First, they showed that the sun's spherical aberration gives rise to a frequently observed property of poor lenses - a caustic, which is a surface of theoretically infinite intensity. In II-9 we will find simpler methods of characterizing this surface, and we will comment on the significance of this to solar physics. Their other major point was that an actual intensity map of the focal region could be formally inverted to yield the radial variation of the density of the sun. The implication of this is startling - the center of the sun might be examined directly, analogous to studies of atomic nuclei by high energy particles.

The present thesis took shape early in 1976, when the author independently put forth these same ideas. The only major difference with the earlier authors is that they did not regard a focal exploration as practical, within the foreseeable future. In contrast, the bulk of the present work is concerned with spacecraft design and performance, and the design of missions for solar study and gravitational wave astronomy. Finally we note, along with Cyranski and Lubkin, that the sun is probably transparent to low energy neutrinos also. However, the author is unaware of any suggested strong cosmic neutrino sources, or of detectors of a mass compatible with a focal mission spacecraft. Should these notions change, a neutrino focal mission could look much like the gravitational wave missions studied here.

2. Missions to the Solar Foci

In Appendix A we shall compute the focal length $\ell(0)$ of the sun for five published solar models, reaching answers in the range $3.36 < \ell(0) < 4.43$ Tm. This range lies between the mean distances of Uranus (2.87 Tm) and Neptune (4.5 Tm). Every remote source of radiation possesses a focus somewhere on a sphere of radius $\ell(0)$ and centered on the sun. The line from a given source, through the center of the sun, and projected through the focus, is called here the axis. We shall show that the focal region (i.e., the region of high relative intensity) is a needle-like shape extending nearly 1 Tm along the axis, but only a few Mm to any side.

There are two main reasons for visiting the foci. First, as we noted above, a detailed intensity map of a focus could yield a density map of the sun, with greatest sensitivity at the core. The possibilities of resolving solar oblateness, and even tesseral features, exist, but will need further study. The other main reason is gravitational wave astronomy. The interest here is in imaging. As we will show in VI-5, source details on the order of 10^{-3} arc seconds may be resolvable in a focal survey. With gravitational radiation yet to be detected at this writing, it is difficult to assess the value of such measurements, but some possibilities are discussed in VI-5.

Laying the groundwork for such missions is the primary intent of this thesis. The first task is straight physics - establishing the locations and structure of the foci. In Chapter II the geometric optics of the sun are worked out from general relativity. This analysis presumably parallels that of Cyranski and Lubkin, but goes beyond their published results in several important respects. The theory is applied to five published solar models in Appendix A. The assumption that the sun lies in, or at least not far from, this range of models is the basis for the selection of several important trajectory and mission design parameters.

I-3

One important mission design constraint is that until a successful focal mission has been flown, the focal length $\ell(0)$ will not be known. As we will show, this implies that a first mission must reach the axis well before the lowest $\ell(0)$ prediction, and then fly straight out until the focus is found. The point at which this purely axial flight begins is called here the "turn point", and on the basis of the results in Appendix A, a turn point distance of 2 Tm is adopted, or about halfway between the distances of Saturn and Uranus.

The shape of the trajectory between the earth ("departure") and the turn point depends strongly on the propulsion methods employed. We will examine the possibilities for chemical propulsion in Chapter III, and ion propulsion in Chapter IV. Various other possibilities and improvements exist, and will be commented on. Several warrant future study.

To keep the thesis within reasonable bounds, it has been necessary to constrain the choice of launch vehicles. Accordingly, we have adopted the ground rule that a focal mission spacecraft shall be designed to fit in a single dedicated Space Shuttle payload. The known mass, volume, acceleration, safety, and orbital constraints will all play important roles in the departure vehicle design, to be synthesized in Chapter III.

To preview the departure vehicle design, a liquid hydrogen and oxygen propulsion system is chosen in III-7, compatible with the main Shuttle propulsion. Existing but slightly modified engines are employed. The propellant tanks of III-8 are novel. They consist of Mylar bags with web supports, and shapes tailored to fit the Shuttle bay. The mass savings over aluminum or titanium are remarkable. The structure of III-9 is also interesting. It consists of a pentagonal truss of beryllium tubing. While this is expensive, the mass saving over the next best material considered (magnesium) is impressive. Also examined were aluminum, titanium, and steel.

The performance of this combination is nearly 14 Mg payload in a parabolic escape orbit at the optimal inclination of 28.5 deg. This is not far from the optimal departure conditions that we will find in IV-10 and 11 for an ion propelled upper stage. However, for a chemically powered upper stage, as discussed in Chapter III, substantial hyperbolic excess is called for, and much lower payloads result.

For an all chemically powered spacecraft, the minimum complexity is offered by a second stage using storable propellants, and fired at the turn point. A rough design for this stage is offered in III-11. Once the turn point distance and post turn point speed have been chosen (the latter from the time of flight considerations of III-2), there is sufficient information for a trajectory optimization. The free variables are the departure and turn point impulse vectors and the earth's orbital position at departure. Constraints consist of departure from earth orbit, final turn point conditions, and celestial mechanics. The quantity optimized is second stage payload. From the detailed results in III-12, we will find that payloads in excess of 1 Mg are possible for foci close to the ecliptic.

Transfer times to the turn point are computed in III-13, where it is found that 4 years is typical for optimal trajectories. Finally, overall mission times are considered in III-14. This depends strongly on $\ell(0)$, but 10 years is a plausible estimate. Trading mission time for payload is a topic for future study.

Several possibilities exist for improving this performance, and are discussed in IV-1. Reasoning that nuclear electric power is unavoidable for outer planetary spacecraft, we are led to consider nuclear-electric propulsion as a replacement for the second stage. For conservatism, the power source is taken to be a scaled-up version of the Voyager spacecraft radioisotope

thermoelectric generator (RTG). With this type of source, power not used is wasted; so it pays to operate the engines at full power continuously, varying only the propellant flow rate. Taking as a performance index the turn point mass less the mass of the RTG, we have a problem in continuous optimal control.

While optimal ion drive trajectory problems are not new, this one is unusual in that the first stage impulse and the RTG mass are regarded as controls, along with the continuous drive acceleration vector. We have also taken account of the slow decay of the RTG. With a half life of 17 years, this effect is important only for the longer transit times to the turn point. However, the mathematical complexity is only slightly increased; and as a computer solution is necessary in any case, the extra realism seemed worthwhile. The full three dimensional problem is a two point boundary value problem of tenth order with two integral path constraints, and 13 boundary conditions, six of which must be initially guessed.

An elaborate computer search, as described in IV-10 and 11, has produced solutions for a wide range of transit times to the turn point, and over the full range of source declinations. An interesting feature of the solutions is that they are divided into separate classes, marked by the number of orbits that are made around the sun, prior to bending outward into a radial trajectory, tangent to the axis at the turn point.

Overall performance (payload at the turn point less the RTG mass) is 5-10 Mg for the range of transfer times (up to 10 years) and source declinations studied (all). Performance increases with transfer time, rapidly at first, but very slowly after about 9 years, and with an upper bound of 13.8 Mg. Payload decreases with source ecliptic declinations; but

quite slowly, in sharp contrast to the all-chemical case. Performance curves and several sample trajectories are pictured in Figures IV-2-14.

While the main research of this thesis is expounded in Chapters II-IV, the author feels that the work would be incomplete without some discussion of actual focal missions, and of gravitational wave detectors appropriate to those missions. Accordingly, two chapters have been added to cover these topics. Partly for brevity and partly because the necessary data is sketchy, these topics are not covered as thoroughly as those in the earlier chapters. However, in spite of analytic short-cuts and approximations, it is often possible to establish the feasibility of various aspects of focal missions, within existing technology.

To begin, "Weber bar" detectors are reviewed in V-2. A combination of low damping single crystal developments, and simple deep space cryogenic techniques are shown to be promising candidates for focal missions. An alternative to Weber bars is the long baseline optical detector, using three or more spacecraft. In V-3-5 such a detector is synthesized from a combination of existing techniques, and optimized for the anticipated frequency of radiation from the Crab Nebula pulsar, 60.427 Hz. Such techniques look very promising for mid-frequency periodic sources such as pulsars; but the anticipated radiation levels are so low that even this may not work. Construction of such a system poses many challenging optical, electronic, and controls design problems.

The final chapter covers a range of more or less disjoint topics; but leading up to a final section (VI-5) on mission design. First (VI-2), there is a review of potential sources of radiation. These range from binary stars, whose radiative properties are precisely calculable within any given theory of gravity, to the more speculative pulsar, globular cluster, and

galactic sources. Ecliptic coordinates for some of the more promising sources are calculated in order to establish the feasibility of reaching their foci by chemical propulsion. (All are reachable by ion propulsion.)

The next section examines the quality of the solar lens. Diffraction, which may turn out to be the worst lens problem, remains an open question at this writing. Various possibilities are mentioned in VI-3. If any of these are eventually proved correct, substantial modification of the geometric optics of Chapter II and Appendix A may be needed. Another lens problem is solar asphericity, which is both a blessing and a curse. The difficulty is that extending the analysis of Chapter II and Appendix A to an aspheric sun is likely to be quite tedious. On the other hand, measurement of the asymmetries of a focal pattern may reveal the actual asphericity of the sun-- a substantial motivation for focal missions. Other problems considered are interference and image size (not really a lens problem). The former appears to matter only at very high radiation frequencies, and the latter only for extended or multiple sources.

Without analysis, it is by no means evident that existing techniques are adequate to navigate beyond the turn point with sufficient accuracy to locate the focus. This is a question which a student in guidance and control finds especially difficult to overlook. In VI-4, navigation based purely on NASA Deep Space Network tracking is examined. After drastic simplification, an easily solvable problem in optimal estimation and control is posed. The main conclusion is that navigation to adequate accuracy is just possible, without expensive long-term operational measures. However, by the addition of an on-board low level accelerometer, substantial improvement is possible. Much further work is needed in this area.

The final section, VI-5, serves to bring it all together. Three classes of focal missions are defined and studied, and mission profiles are developed for each. The three classes are the first mission for determining the focal length $\ell(0)$, solar missions for determining the internal structure of the sun, and astronomical missions for examining the spatial structure of interesting sources. In all cases it is assumed that gravitational radiation from the intended source has somehow been detected, prior to mission planning, that the intended detector can sense this source even without focal enhancement; and that precise optical or radio astrometric source data is available.

For each mission type, a rationale is developed, feasible detector-propulsion combinations are found, and the search, acquisition, and focal study phases are delineated. A variety of complex trajectory design and optimization problems are identified for future study. In general, the goal of each is to maximize the scientific yield, while not letting the overall mission time become excessive. The mission design for each type will be an extremely challenging task.

3. Directions for Future Research

If, at the completion of this long thesis, one thing stands out, it is how much there is left to do. From the many speculations and open questions in the text, a number have been culled for discussion here. The questions are grouped under five headings -- physics, transfer orbit studies, detectors, mission optimization, and navigation. We will take them up in this order.

By almost any reckoning, the outstanding open physics question is diffraction. The present treatment (and that of Cyranski and Lubkin) corresponds to geometric optics. That this may be inadequate for low frequency sources is commented on in VI-3; but no suggested line of attack is offered. If focal research is to be pursued, understanding diffraction should rate first priority.

A group of studies bearing on the first and solar research missions derive from the reference solar model, as discussed in VI-5. The present work deals with five published models, but for most illustrative calculations, relies on the Ezer-Cameron model[39]. As this 13 year-old model could probably be improved on today, a study should be undertaken to derive a new spherically symmetric model -- the reference solar model, as it is called here. From this model, and its more plausible competitors, the likelihood density $p(\ell)$ that the focus will be found at the distance ℓ should be estimated. This is essential to the planning of the first mission.

Another aid to planning is the full three dimensional solar model. In this the solar density is described by some sort of expansion, whose co-

efficients we hope to determine from solar missions. To plan an optimal focal exploration, we need to weight the relative scientific importance of these coefficients. This point is discussed further below and in VI-5.

As gravitational theories differ greatly in their predictions about gravitational radiation, the first unequivocal detection is likely to weed out several currently viable theories. An interesting possibility is that an examination of focal structure may subject the survivors to a still more discriminating test. In fact, the detection of any focal structure would rule out a few theories in which gravitational waves, unlike light, propagate in a flat background metric. Except for these, the present Chapter II could be redone within other theories, and the gain picture constructed for each, as in Appendix A, using the reference solar model. Substantial differences in the pictures may result. To avoid unnecessary work, it is probably best to hold this study until after the first detection has culled the field.

A final physics effort of great value to focal studies will be an improved source survey, giving new estimates on the local signal strengths, frequency spectra, and polarizations of plausible sources. It is recognized that an adequate survey may not be possible in the near future. Of course, efforts along this line are certain, with or without further focal studies.

Turning to engineering studies, several possibilities for improving transfer orbit performance over the chemically and electrically powered vehicles examined here are suggested in the text. Perhaps first amongst these is planetary gravity assists. This would greatly increase chemical capabilities, and may be helpful for ion propelled vehicles as well. The problem here is that there are many ways to get single or multiple planetary

assists, and examining any one is much harder than the corresponding problem treated here. Thus, total sky studies, as done here, may not be feasible, and interesting missions would have to be studied individually. A related possibility is the addition of a third stage to the all-chemical vehicle. In combination with a gravity assist, performance might be considerably extended.

Improvements over the electrical propulsion system of Chapter IV are possible. First, the ion engine efficiency, taken as 50% here, should be reviewed. Also, a better RTG than that flown on Voyager is now becoming available, and for missions more than a decade off, a fission reactor power supply may be feasible. However, none of these improvements require any qualitative change in the theory of Chapter IV. Ion drives could be extended in another way if some or all of the first stage impulse were achieved electrically. Much new theory would be required. Finally, there are some holes in the present theory. The trajectories computed here are certainly local optima; but the global properties remain somewhat conjectural. Also, the division of trajectories into overlapping classes suggests that conjugate points may exist. A deeper look at these questions may lead to better understanding.

A really intriguing possibility is to combine the ion drive with a solar sail, as discussed in IV-1. Since a nuclear power source is needed in any case, a sail by itself is not very attractive. However, outer planet solar sail trajectories look remarkably similar to the ion drive trajectories found here. Thus, a combination might yield substantial improvements in performance. Considerable preliminary design work is needed in order to construct a spacecraft model, suitable for optimization, as in Chapters III and IV.

A final transfer problem needing study is non-optimal departures. In III-6 it is shown that optimal departure for a chemical mission is possible only over a limited band of source latitudes and longitudes. "Optimal" here means that the departure impulse can be lined up with the Shuttle orbital velocity. Missions to foci lying outside this band have not been studied. For ion drives, the greater control flexibility makes it plausible that all departures will be optimal, regardless of source location. However, this point needs proof.

Returning to physics, the design of gravitational wave detectors for use in space has a long way to go. Efforts in this direction, on both compact and long wave detectors are likely, regardless of the future of focal studies. Thus many of the outstanding propulsion, navigation, and control problems identified in Chapters V and VI may receive attention as well. In particular, the effect of ion propulsion process noise on the detector will have to be addressed.

Some detector design studies are unique to focal missions. One of these is to find the range of applicability of the more sensitive long wave detectors in the confined neighborhoods of the foci. This is needed in the decisions on what types of detectors are suitable for each of the missions examined in VI-5. As for compact detectors, in V-2 it is shown that cryogenic techniques are uniquely suited to focal missions. However, a study is needed to see how far we can go below the 20K passive radiator examined there, without excessive power, weight, and complexity.

The attempts at mission synthesis in VI-5 have identified a host of unsolved problems. Perhaps the first of these is the selection of the turn point radius r_T for the first mission. This should be as far out as possible to maximize performance; but not so far as to risk bypassing the focus. A

plausible compromise might be where the probability of the focus lying within r_T is around 0.01; as determined from the likelihood density $p(\ell)$ discussed above. A less conservative choice may be necessary if chemical propulsion then yields too little payload. The selection of the post-turn point speed v_T for the first mission is another difficult optimization, for which an approach is suggested in VI-5. That both r_T and v_T have strong effects on the payload, at least for chemical propulsion, can be seen in Figure III-9. Similar studies for ion propulsion have not yet been carried out.

The selections of r_T and v_T for solar and astronomical missions are different kinds of problems, since $\ell(0)$ will then be known. Some discussion is given in VI-5. In these cases the selection of r_T and v_T become part of the design of the acquisition sequence. This is the portion of the trajectory following the first detection of focal effects, whose purpose is the precise location of the focus. The optimization is yet to be properly formulated, and may be as difficult as that attempted here.

The next set of studies is needed to design the focal tour; i.e., the final parts of the trajectories, following acquisition, on the first and solar missions. This starts from the reference solar model discussed above. First, the gain contour picture of Appendix A should be repeated for this model. Diffraction effects should be well understood before attempting this, as there may be important source frequency effects. The scalar gain field so obtained will possess scalar field derivatives with respect to each coefficient of the full solar model, including aspheric terms. This will aid the mission design in that if the measurement of a particular coefficient is desired, the tour should include regions where that derivative field is large.

More precisely, the process of extracting the actual solar model coefficients from a sequence of focal intensity measurements is greatly

facilitated by the gain field derivatives with respect to those coefficients.
In addition, from the given scientific value of each coefficient, as dis-
cussed above, a performance index involving the relative desired accuracies,
along with propellant usage and mission time, could be formulated. This
would provide a tool with which different types of focal tours could be
compared. Once again, the gain derivatives would play a central role.
Whether a full optimization, such as is attempted in Chapter IV, is possible
remains to be seen. Both the algebraic and the computational difficulties
look immense. Some further discussion of this is given in VI-5.

The last set of mission studies concern the design of astronomical
missions. These are missions in which a fleet of ion powered detectors
explore the focal image of a source. It is shown in VI-2 and VI-5 that good
candidates for these missions are likely to be irregular burst sources, with
fairly frequent bursts. This complicates the acquisition process in that the
signal amplitudes must be compared between the members of the fleet to locate
the axis. Unless it becomes possible to make absolute internal sensitivity
calibrations of the instruments, it will be necessary to design fleet maneuvers
in such a way that cross-calibration of the detector ensemble is possible
during acquisition, after several detected bursts. The design of these fleet
cross-calibration maneuvers is still another challenging problem.

On arriving at the focus the fleet comes to a near halt, and
attempts to map the focal image. Here, distortions of the focal intensity
picture, previously known from solar missions, are interpreted as multiple
or complex sources. Potential sources, such as the center of our galaxy,
might show interesting structure in this way. Once again we face the question
of how to arrange the fleet. It is much more complicated here in that pre-
liminary indications of structure would lead to redeployment in order to

improve the resolution. Thus we are faced with a feedback control problem, whose initial conditions are themselves the product of an optimization.

A unique feature of astronomical missions is that there are often multiple targets. The nucleus and "jet" of M87, and the bevy of quasars "near" M82, are suggested targets in VI-2 and VI-5. After examining the focal image of one source component, the fleet would jump to another. The economics of this are discussed in VI-5. This time the new focus would be approached laterally, which may modify the acquisition process. Further complexities arise in that the new source component might not emit much gravitational radiation. Thus a lengthy and possibly unfruitful search may be needed. The design of the joint fleet search pattern is yet another difficult task.

Finally, in connection with multiple sources, there is the question of the order of visits. If the criterion is the minimization of total jump time, then we are faced with a variation of the well-known "traveling salesman" problem. The jump sequence problem becomes a bit more complex if we add a priority list; i.e., penalize the total time to reach some source foci more than others. In either case, for N source components, there are N! distinct tours, and direct comparison of all these becomes very expensive for $N \gtrsim 8$. Methods for weeding out grossly uncompetitive tours would be very desirable.

The last group of needed studies concerns navigation. In VI-4, a highly simplified study shows that the NASA Deep Space Network (DSN), using differenced range techniques, is just capable of navigation following the turn point. Two to three modest antennas, transmitters, and receivers would be needed nearly full time over several years. It is also found that the DSN

burden can be substantially relieved by incorporating an on-board accelerometer. Various studies are needed to refine these conclusions.

First, we need a study of the economics of the various ways in which the DSN could support focal missions. The parameters include ground and spacecraft equipments, frequencies, system calibration techniques, and operating modes such as differenced range, doppler, and differential long baseline interferometry. The output would be tracking residual biases, variances, and correlation times versus cost for the more competitive combinations.

The DSN study provides most of the inputs needed to redo the optimal estimation and control study in VI-4. This time the approximations of continuous information, independence of the cross axes, and independence of the controls and process noise could be avoided, or at least better justified. For this, a believable model of process noise is essential. The output here would be the estimation, control, and navigation error covariances for each set of input assumptions. In a variation of this study, an accelerometer would be added with various levels of bias and measurement noise statistics. When all this is completed, we should know what it will cost to meet the navigational requirements under each of several options.

Some alternative measurements also need study. One of these would be an optical instrument capable of determining planetary positions relative to the stars. While the necessary accuracy (of order 10^{-7} radian) will be difficult to achieve, there is an inherent advantage over an accelerometer -- gravitational perturbations are not ignored. The study should be aimed primarily at finding the instrument accuracy required to relax the DSN requirements significantly, with no accelerometer.

As a final alternative we mention the possibility of interplanetary navigational beacons. A set of at least three of these, synchronously transponding an earth master station, would amount to a three dimensional Loran system. While the cost of such a system would be high, its accuracy would greatly surpass that of the DSN, even if augmented with a first rate accelerometer. Of course, this is not really a topic for focal mission studies -- many proposed interplanetary missions would benefit from a beacon system.

CHAPTER II

SOLAR FOCI

1. Static Radial Fields

Many of the gross features of solar focussing can be found from a solar model consisting of a static, radially symmetric ball of perfect fluid. Following the treatments of Weinberg[94] and MTW[66] we will derive the deflection of null geodesics in such a field.

We begin with the line element. From MTW, Eq. 23.7, the most general static, radially symmetric line element can be written as:

$$ds^2 = - e^{2\Phi}dt^2 + e^{2\Lambda}dr^2 + r^2d\theta^2 + r^2s^2\theta d\phi^2 \tag{1}$$

where $\Phi = \Phi(r)$ and $\Lambda = \Lambda(r)$. Weinberg[94] obtains an equivalent form, Eq. 8.1.6. In (1), the coordinate time t is that measured by a remote observer $(r \to \infty)$ in flat spacetime. The coordinate r is a label for the radius such that a great circle at constant r has proper circumference $2\pi r$, or a sphere has proper area $4\pi r^2$. The coordinates ϕ and θ are the usual spherical coordinates (longitude and colatitude). Together, these are Schwartzchild coordinates.

It is easy to see that a free trajectory in such a field must be contained in a plane passing through $r = 0$. Thus, without loss of generality, we can take $\theta = \pi/2$, and reduce (1) to:

$$ds^2 = g_{ij}dx^idx^j = - e^{2\Phi}dt^2 + e^{2\Lambda}dr^2 + r^2d\phi^2 \tag{2}$$

where i and j range over t, r, and ϕ.

Any free falling particle follows a geodesic. This means that the trajectory must obey:

$$\frac{d^2x^i}{d\psi^2} + \Gamma^i_{jk} \frac{dx^j}{d\psi} \frac{dx^k}{d\psi} = 0 \tag{3}$$

.The affine connections or Christoffel symbols are:

$$\Gamma^i_{jk} = \frac{1}{2} g^{i\ell}(g_{\ell j,k} + g_{\ell k,j} - g_{jk,\ell}) \tag{4}$$

while ψ is some parameter describing the trajectory. From (4), after noting that the g_{ij} depend only on r, we can easily work out all the Γ^i_{jk}. Those not vanishing are:

$$
\begin{rcases}
\Gamma^t_{rt} = \Gamma^t_{tr} = \Phi'(r) \\[2mm]
\Gamma^r_{tt} = \Phi'(r)e^{2(\Phi-\Lambda)} \\[2mm]
\Gamma^r_{rr} = \Lambda'(r) \\[2mm]
\Gamma^r_{\phi\phi} = - re^{-2\Lambda} \\[2mm]
\Gamma^\phi_{r\phi} = \Gamma^\phi_{\phi r} = 1/r
\end{rcases} \tag{5}
$$

Substituting (5) into (3) gives these path equations:

$$\frac{d^2t}{d\psi^2} + 2 \Phi' \frac{dt}{d\psi} \frac{dr}{d\psi} = 0 \tag{6}$$

$$\frac{d^2r}{d\psi^2} + \Phi' e^{2(\Phi-\Lambda)}\left(\frac{dt}{d\psi}\right)^2 + \Lambda'\left(\frac{dr}{d\psi}\right)^2 - re^{-2\Lambda}\left(\frac{d\phi}{d\psi}\right)^2 = 0 \tag{7}$$

$$\frac{d^2\phi}{d\psi^2} + \frac{2}{r} \frac{dr}{d\psi} \frac{d\phi}{d\psi} = 0 \tag{8}$$

Of these, first integrals of (6) and (8) are immediately available:

$$e^{2\Phi} \frac{dt}{d\psi} = K_1 \qquad (9)$$

$$r^2 \frac{d\phi}{d\psi} = K_2 \qquad (10)$$

Putting these into (7) yields another integral:

$$e^{2\Lambda}\left(\frac{dr}{d\psi}\right)^2 - K_1^2 e^{-2\Phi} + K_2^2 r^{-2} = K_3 \qquad (11)$$

So far, we have made no use of the fact that we are dealing with null geodesics. In such cases the line element vanishes. Thus, putting (9-11) into (2) we get:

$$K_3 = \left(\frac{ds}{d\psi}\right)^2 = 0 \qquad (12)$$

We will also find it convenient to identify the point of closest approach or perihelion. Here we define $r = r_0$ and $\Phi(r_0) = \Phi_0$. Also, we can arrange our coordinates so that $\phi(r_0) = 0$, and increasing. Since r is a minimum here, $dr/d\psi = 0$, and from (11) and (12):

$$K_1 r_0 = K_2 e^{\Phi_0} \qquad (13)$$

Next, we use (10) to eliminate ψ from (11). With the help of (12) and (13) this gives:

$$\left(\frac{1}{r}\frac{dr}{d\phi}\right)^2 = e^{-2\Lambda}\left[\left(\frac{r}{r_0}\right)^2 e^{2(\Phi_0 - \Phi)} - 1\right] \qquad (14)$$

This is the differential equation of the path in Schwartzchild coordinates. It can be integrated directly as soon as the metric functions Φ and Λ are specified.

2. Deflection

The halves of the trajectory on each side of perihelion will be symmetric; so to compute the deflection we can restrict attention to the region $r_0 \leq r \leq \infty$; $0 \leq \phi \leq \frac{\pi}{2} + \varepsilon$; where ε is a small angle representing half the deflection. If (14) is integrated directly, then $\varepsilon = \phi(\infty) - \frac{\pi}{2}$. Thus, a numerical integration would require extreme precision. Two methods have been found to circumvent this.

The most obvious is to introduce a new variable σ by:

$$c\sigma = r_0/r \qquad (15)$$

This has the range $0 \leq \sigma \leq \pi/2$, and increases monotonically with r. We can then define

$$\varepsilon = \phi - \sigma \qquad (16)$$

This gives ε the correct final value, and we will show that it is everywhere small for solar trajectories.

From (15) we have:

$$\tan \sigma \, d\sigma = \frac{dr}{r} \qquad (17)$$

and putting (15-17) into (14):

$$e^{2\Lambda}\tan^2\sigma\left(\frac{d\sigma}{d\varepsilon+d\sigma}\right)^2 = e^{2(\Phi_0-\Phi)}\sec^2\sigma - 1 \qquad (18)$$

which can be rewritten as:

$$\left[e^{2(\Phi_0-\Phi)} - 1 + s^2\sigma\right]\left(1 + \frac{d\varepsilon}{d\sigma}\right)^2 = e^{2\Lambda}s^2\sigma \qquad (19)$$

Anticipating that Φ and Λ are everywhere tiny for the sun, we see that we can subtract $s^2\sigma$ from both sides and be left with a relation between tiny quantities. Unfortunately the result is of second degree in $d\epsilon/d\sigma$, and requires either linearization, or iteration, or both for its solution. It is clear though, that ϵ is everywhere tiny.

The other method avoids the second degree problem by a curious finesse. Suppose we form the derivative in (14) by replacing $d\sigma$ by $d\phi$ in (17). This requires a new variable λ:

$$\tan \lambda \, d\phi \equiv \frac{dr}{r} = d(\ln r) \tag{20}$$

It is evident from (20) that like σ, λ has the range $0 \le \lambda \le \pi/2$. Putting (20) into (14) yields:

$$e^{2\Lambda}\tan^2\lambda = \left(\frac{r}{r_0}\right)^2 e^{2(\Phi_0-\Phi)} - 1 \tag{21}$$

This plays the role of (15) in the previous method in that it allows λ to be calculated directly from r, once Φ and Λ are given. We also need to differentiate (21), After a little algebra we get:

$$F(r,\lambda) \equiv r \tan \lambda \, \frac{d\lambda}{dr} = [1-r\Phi'(r)] \, [1-(1-e^{-2\Lambda})c^2\lambda]-r\Lambda'(r)s^2\lambda \tag{22}$$

This time to compute the deflection we introduce the new variable δ:

$$\delta = \phi - \lambda \tag{23}$$

This is similar to ϵ in (16) and, in fact, $\delta(\infty) = \epsilon(\infty) =$ half the final deflection, We shall actually call δ the deflection, but note that the total bending angle as seen by a remote observer is $2\delta(\infty)$. Putting (23) into (20) and (22) gives:

$$1 + \frac{d\delta}{d\lambda} = \frac{1}{F} \qquad (24)$$

so that

$$F \frac{d\delta}{d\lambda} = 1 - F = r \Lambda's^2\lambda + r\Phi' + (1-r\Phi')(1-e^{-2\Lambda})c^2\lambda \qquad (25)$$

This is it. The expression on the right is seen to be tiny for the sun, so $F \doteq 1$, and δ is tiny. The system (21) and (25) can be integrated directly, without approximations or excessive digital precision.

3. The Perfect Fluid Assumption

To make further progress we need to relate the functions $\Phi(r)$ and $\Lambda(r)$ to existing solar models. To this end we will make the further assumption that the sun is a perfect fluid. That is, for our purposes, the contributions of shear stress (viscosity) and energy transport to the stress-energy tensor can be neglected. For a justification of this, based on solar models see Appendix A.

A full relativistic treatment of a static, radially symmetric star, composed of a perfect fluid is given by MTW[66] in Sections 23.4 and 23.5. It is enough to state the main results:

$$e^{-2\Lambda} = 1 - \frac{2m(r)}{r} \qquad \text{(MTW 23.18)} \qquad (26)$$

where

$$m(r) = 4\pi \int_0^r x^2\rho(x)dx \qquad \text{(MTW 23.19 et. seq.)} \qquad (27)$$

Here $\rho(r)$ is the mass-energy density, and $m(r)$ is the total mass-energy

within radius r, which is the Schwartzchild coordinate radius used above. To be consistent with MTW[66], gravitational units are adopted, in which G = c = 1; so that m is a length, and ρ is (length)$^{-2}$.

The other needed result is:

$$\Phi'(r) = (\frac{m}{r^2} + 4\pi r p)e^{2\Lambda} \qquad \text{(MTW 23.21)} \qquad (28)$$

In this p is the pressure in units (length)$^{-2}$. It may seem unnecessarily fussy to carry the pressure term here; but it is little trouble, and we are spared justifying its neglect. For each given solar model in Appendix A, (28) is integrated numerically to give Φ, as required in (21).

To make use of these we first differentiate (26), and combine with (27). This gives:

$$\Lambda'e^{-2\Lambda} = \frac{m'(r)}{r} - \frac{m(r)}{r^2} = 4\pi r\rho(r) - \frac{m(r)}{r^2} \qquad (29)$$

Then putting (28) and (29) into (25) leads to:

$$F\frac{d\delta}{d\lambda} = 1 - F = 4\pi r^2 e^{2\Lambda}(\rho + p)s^2\lambda + (\frac{3m}{r} + 4\pi r^2 p)c^2\lambda \qquad (30)$$

The system (21), (26), (28), and (30) is now readily integrated for any given solar model.

4. Non-dimensional Equations

One further convenience in numerical work is to eliminate dimensions in such a way as to make the more important variables ≈1. To do this we will introduce the solar "radius" r_s as that value where $\rho(r) = p(r) = 0$

for $r > r_s$, and not otherwise. Then from (27) the mass of the sun (total mass-energy) is $m_s = m(r_s)$, and $m(r) = m_s$ for $r \geq r_s$. We will also introduce an "average" density:

$$\bar{\rho} = \frac{3m_s}{4\pi r_s^3} \tag{31}$$

The quotes are because this is a purely formal definition - $\frac{4}{3}\pi r_s^3$ is not a covariant volume. The values in MTW[66] are:

$$r_s = 6.9598 \times 10^8 m$$

$$m_s = 1.989 \times 10^{30} kg = 1476.64 \ m$$

$$\bar{\rho} = 1.4085 \ gm/cm^3 = 1.04567 \times 10^{-24} m^{-2}$$

Slightly different values are adopted in Appendix A to conform to specific solar models.

Using these definitions we get these normalized variables:

$$R = \frac{r}{r_s}; \quad M = \frac{m}{m_s}; \quad W = \frac{\rho}{\bar{\rho}}; \quad P = \frac{p}{\bar{\rho}} \tag{32}$$

We also define

$$\gamma = m_s/r_s = 2.12167 \times 10^{-6} \tag{33}$$

With these the system (26), (28), (21), and (30) becomes:

$$e^{-2\Lambda} = 1 - 2\gamma \frac{M}{R} \tag{34}$$

$$\Phi'(R) = \gamma e^{2\Lambda} \left(\frac{M}{R^2} + 3RP \right) \tag{35}$$

II-8

$$e^{2\Lambda}\tan^2\lambda = \left(\frac{R}{R_0}\right)^2 e^{2(\Phi_0-\Phi)} - 1 \tag{36}$$

$$F\frac{d\delta}{d\lambda} = 1 - F = 3\gamma[R^2 e^{2\Lambda}(W+P)s^2\lambda + (\frac{M}{R} + 3R^2P)c^2\lambda] \tag{37}$$

where

$$R_0 = r_0/r_s \tag{38}$$

To sum up, (34-37) is a system for computing the deflection of a null geodesic by the sun requiring only the evaluation of integrals. It assumes only Einstein general relativity, and that the sun is a static, spherically symmetric ball of perfect fluid. There are no approximations.

5.   The Exterior Deflection

The portion of the deflection δ occurring outside the sun ($R \geq 1$) is independent of the solar model, apart from small differences in γ. Also (34-37) take particularly simple forms, and we will carry out the integration analytically to $O(\gamma^2)$. This development is not really essential to the first order theory that follows; but it is interesting, and it may point the way to future extensions of the theory to higher order. By definition:

$$W = P = 0; \quad M = 1 \tag{39}$$

Thus from (34):

$$e^{-2\Lambda} = 1 - \frac{2\gamma}{R} \tag{40}$$

This allows us to integrate (35):

$$\Phi(R) = \gamma \int_{\infty}^{R} e^{2\Lambda(x)} \frac{dx}{x^2} = \gamma \int_{\infty}^{R} \frac{dx}{x(x-2\gamma)}$$

$$= \frac{1}{2} \ln\left(1 - \frac{2\gamma}{R}\right) = -\Lambda(R) \tag{41}$$

in which, for simplicity, we have made the conventional choice $\Phi(\infty) = 0$. With this, (36) simplifies to

$$\sec^2\lambda = \left(\frac{R}{R_0}\right)^2 e^{2\Phi_0} + \frac{2\gamma}{R} \tag{42}$$

and finally, (37) becomes

$$F \frac{d\delta}{d\lambda} = 1 - F = \frac{3\gamma}{R} c^2\lambda \tag{43}$$

At this point we need to assume that Φ_0 is $O(r)$. In Section 7, this will be demonstrated for simple polynomial density models of the sun, and in Appendix A it will be confirmed by direct numerical integration of (34) and (35) for the models treated there. Only a model departing greatly from these "conventional" models, such as a collapsed core, or a central black hole, would violate our assumption. Applying this to (42):

$$\sec^2\lambda = \left(\frac{R}{R_0}\right)^2\left(1 + 2\Phi_0 + \frac{2\gamma R_0^2}{R^3}\right) + O(\gamma^2) \tag{44}$$

from which

$$c\lambda = \frac{R_0}{R}\left(1 - \Phi_0 - \frac{\gamma}{R_0} c^3\lambda\right) + O(\gamma^2) \tag{45}$$

With this we can eliminate R from (43):

$$\frac{d\delta}{d\lambda} = \frac{3\gamma}{R} \left(1 + \frac{3\gamma}{R} c^2\lambda\right) c^2\lambda + O(\gamma^3)$$

$$= \frac{3\gamma}{R_o} \left(1 + \Phi_o + \frac{4\gamma}{R_o} c^3\lambda\right) c^3\lambda + O(\gamma^3) \qquad (46)$$

There is no difficulty with any of these expressions as $R_o \to 0$, as $c\lambda \lesssim R_o$.

As one application of this we will consider the starlight bending problem, first solved by Einstein in 1916[37], using an equivalence principle argument. In this case $R_o \geq 1$. First consider the $O(\gamma)$ term:

$$\delta(R) = \frac{3\gamma}{R_o} \int_0^\lambda c^3\theta d\theta = \frac{\gamma}{R_o} (2 + c^2\lambda)s\lambda \qquad (47)$$

For $R = \infty$, this gives Einstein's well known result for the total bending, which at grazing incidence ($R_o = 1$) is:

$$2\delta(\infty) = 4\gamma = 8.48668 \times 10^{-6} \text{ rad} = 1.7505 \text{ sec}$$

The effect of finite R is surprisingly small. Let

$$G(\lambda) = \left(1 + \frac{1}{2} c^2\lambda\right)s\lambda = 1 - \frac{3}{8} c^4\lambda - \frac{1}{8} c^6\lambda - \cdots \qquad (48)$$

Now at the position of the earth

$$R = \frac{r_e}{r_s} = \frac{1.495985 \times 10^{11}}{6.9598 \times 10^8} = 214.946$$

and the correction factor is essentially

$$\frac{3}{8} c^4\lambda = \frac{3}{8R^4} = 1.76 \times 10^{-10}$$

In fact, for $R = 3R_0$, $\delta(R)$ is already within $1/2\%$ of $\delta(\infty)$. The message here is that unless we are very close to the sun ($R < 10$ say), then $\delta(R) = \delta(\infty)$ for practical purposes.

To carry this result to $O(\gamma^2)$, we can from (41) put $\Phi_0 = -\gamma/R_0$. Then (46) leads to

$$\delta(R) \doteq \delta(\infty) = \frac{2\gamma}{R_0} \left(1 - \frac{\gamma}{R_0}\right) + \frac{12\gamma^2}{R_0^2} \int_0^{\pi/2} c^6 \lambda \, d\lambda$$

$$= \frac{2\gamma}{R_0} \left[1 + \frac{\gamma}{R_0} \left(\frac{15\pi}{16} - 1\right)\right] \tag{49}$$

This time the correction factor at grazing incidence is 4.127×10^{-6}, which amounts to 7.22×10^{-6} sec. This correction is bigger than that for finite R, for reasonably large R, but still negligible for practical purposes.

Turning now to rays with $R_0 \lesssim 1$, the external deflection is found by integrating (46):

$$\delta_e(\infty) = \frac{3\gamma}{R_0} (1 + \Phi_0) \int_{\lambda_1}^{\pi/2} c^3 \lambda \, d\lambda + \frac{12\gamma^2}{R_0^2} \int_{\lambda_1}^{\pi/2} c^6 \lambda \, d\lambda + O(\gamma^3) \tag{50}$$

where from (45):

$$c\lambda_1 = c\lambda(1) = R_0(1 - \Phi_0 - \gamma R_0^2) + O(\gamma^2) \tag{51}$$

In the work that follows, we shall be primarily interested in nearly central rays; i.e. those for which $R_0 < 0.15$. In this case, the integrals in (50) are quite small, and are best treated by an expansion. Setting

$$\theta = \frac{\pi}{2} - \lambda \tag{52}$$

the $c^3\lambda$ integral becomes:

$$\int_{\lambda_1}^{\pi/2} c^3 \lambda d\lambda = \int_0^\theta s^3 \theta d\theta = \int_{\theta=0}^{\theta=\theta_1} \frac{s^3\theta}{c\theta} d(s\theta) = \int_0^{s\theta_1} \frac{x^3 dx}{\sqrt{1-x^2}}$$

$$= \int_0^{c\lambda_1} x^3(1 + \frac{1}{2} x^2 + \frac{3}{8} x^4 + \cdots)dx$$

$$= \frac{1}{4} c^4 \lambda_1 + \frac{1}{12} c^6 \lambda_1 + \frac{3}{64} c^8 \lambda_1 + \cdots \tag{53}$$

Similarly

$$\int_{\lambda_1}^{\pi/2} c^6 \lambda d\lambda = \frac{1}{7} c^7 \lambda_1 + \frac{1}{18} c^9 \lambda_1 + \cdots \tag{54}$$

It is now evident that for central rays, the external deflection is of $O(\gamma R_0^3)$, and is only a minor contribution to the whole. Since $R_0^7 < \gamma$, we are justified in dropping terms of $O(\gamma R_0^8)$. Thus, combining (50-4) we get

$$\delta_e(\infty) = \frac{3}{4} \gamma R_0^3(1 + \frac{1}{3} R_0^2 + \frac{3}{16} R_0^4) + O(\gamma R_0^9) \tag{55}$$

It is interesting that even to this precision, the structural details of the sun (Φ_0) do not appear. This is because the exit angle from the sun, λ_1, depends only weakly on Φ_0 (see (51)).

Again we need to consider the error due to finite R. From (50) and (53), the fractional error from this source is essentially R^{-4}. Thus, from (55), the error in δ_e is of $O(\gamma^3)$ if $R_0 < 0.15$ and $R > 154$. The latter corresponds to $r > 107$ Gm, or just about the orbit of Venus.

6. Analysis to First Order

While the system (34-7) can be integrated directly for any given R_0 and model of the sun, the process is very tedious, and the results may be hard to interpret physically. What we would prefer is an analytic form $\delta = \delta(R_0)$, whose terms would be physically understandable, and whose parameters could be obtained directly from the solar model. This goal will be shown to be attainable if we restrict our analysis to $O(\gamma)$, and consider only nearly central rays ($R_0 \lesssim 0.15$ say).

To do this we must assume that Λ, P, and Φ are all of $O(\gamma)$ everywhere. This is certainly valid for the solar models used in Appendix A; however, less "conventional" models, particularly those involving a collapsed core or central black hole, have been entertained. Such models are hereafter excluded, at the risk that someday the evidence might invalidate our main conclusions.

Once the $O(\gamma^2)$ terms are deleted, (34-7) reduce to:

$$c\lambda = c\sigma = R_0/R \tag{56}$$

$$\frac{d\delta}{d\sigma} = 3\gamma(R^2 W s^2\sigma + \frac{M}{R} c^2\sigma) \tag{57}$$

Also, from (27), (31), and (32):

$$M(R) = 3 \int_0^R x^2 W(x)\,dx \tag{58}$$

In what follows we shall be attempting to isolate the effects of the parameter R_0. Because of this, the intermediate variable σ is no longer useful.

Eliminating it, and integrating, the system (56-7) becomes:

$$\delta(R_0) = 3\gamma R_0 \int_{R_0}^{\infty} \left[(R^2 - R_0^2)^{\frac{1}{2}} \frac{W(R)}{R} + R_0^2(R^2 - R_0^2)^{-\frac{1}{2}} \frac{M(R)}{R^4}\right]dR \qquad (59)$$

where we have assumed that we are observing the deflected ray at a sufficiently great distance that the result is indistinguishable from the result at infinity to $O(\gamma)$. (See last section.)

The last term in (59) may be integrated by parts:

$$R_0^2 \int_{R_0}^{\infty} (R^2 - R_0^2)^{-\frac{1}{2}} R^{-4} M(R) dR = \frac{1}{3R_0^2} \int_{R=R_0}^{R=\infty} M(R) d\left[R^{-3}(R^2 - R_0^2)^{\frac{1}{2}}(2R^2 + R_0^2)\right]$$

$$= \frac{1}{R_0^2}\left[\frac{2}{3} - \int_{R_0}^{\infty}(R^2 - R_0^2)^{\frac{1}{2}}(2R^2 + R_0^2)\frac{W(R)}{R}dR\right]$$
$$\qquad (60)$$

in which (58) has been used in the last step. With this (59) becomes:

$$\delta(R_0) = \frac{2\gamma}{R_0} - \frac{6\gamma}{R_0}\int_{R_0}^{1}(R^2 - R_0^2)^{\frac{1}{2}} RW(R) dr \qquad (61)$$

The upper limit has been reduced because $W(R) = 0$ for $R > 1$. That the two terms of (57) could be combined by an integration by parts was first noticed by Professor C. M. Will, without benefit of the suggestive arrangement (56-8). This has greatly compressed the analytic drudgery that follows.

To make further progress it has been found convenient to expand $W(R)$:

$$W(R) = W(0) + RW'(0) + \frac{1}{2}R^2 W''(0) + \cdots = \sum_{j=0}^{\infty}\frac{1}{j!}W^{(j)}(0)R^j \qquad (62)$$

This implies that $W(R)$ is analytic on the range $0 \leq R \leq 1$. As we have excluded black holes, the only physically plausible violation of this would be a stratified sun, involving jump discontinuities in $W(R)$. This does not seem very likely; but even then we could closely approximate $W(R)$ by an analytic function of sufficient complexity that the integral (61) would be only trivially affected.

One reason for this expansion is that for a gas or plasma sphere in hydrostatic equilibrium $W'(0) = 0$. To see this we note that the pressure gradient is proportional to the local gravitational acceleration, so that

$$\frac{dP}{dR} = K \frac{M(R)}{R^2} \tag{63}$$

Thus for very small R we get

$$P = P(0) - K_1 R^2 \tag{64}$$

Thus $P'(0) = 0$, and for any plausible equation of state $W'(0) = 0$, provided $T'(0) = 0$, where T is the temperature. The latter must hold as $T'(0) \neq 0$ implies a temperature cusp, which can't be stable in a static gas or plasma. For a more elaborate discussion of this, see for instance Reference 78.

Now applying (62) to (61):

$$\delta(R_0) = \frac{2\gamma}{R_0} - \frac{6\gamma}{R_0} \sum_{j=0}^{\infty} \frac{1}{j!} W^{(j)}(0) G_j(R_0) \tag{65}$$

where

$$G_j(R_0) \equiv \int_{R_0}^{1} \left(1 - \frac{R_0^2}{R^2}\right)^{\frac{1}{2}} R^{j+2} dR \tag{66}$$

A convenient way to treat these integrals as a class is to expand the radical:

$$\left(1 - \frac{R_0^2}{R^2}\right)^{\frac{1}{2}} = 1 - \sum_{i=1}^{\infty} C_i \left(\frac{R_0}{R}\right)^{2i} \tag{67}$$

where

$$C_i = \frac{i(2i-2)!}{2^{2i-1}(i!)^2} = \frac{1 \cdot 1 \cdot 3 \cdot 5 \cdot 7 \cdots (2i-3)}{2 \cdot 4 \cdot 6 \cdots (2i)} \tag{68}$$

This is valid over the closed range $R \geq R_0$. Thus (66) becomes:

$$G_j(R_0) = \int_{R_0}^{1} R^{j+2} dR - \sum_{i=1}^{\infty} C_i R_0^{2i} \int_{R_0}^{1} R^{j+2-2i} dR$$

$$= \frac{1}{j+3}(1 - R_0^{j+3}) - \sum_{i=1}^{\infty} \frac{C_i}{j+3-2i}(R_0^{2i} - R_0^{j+3}) \tag{69}$$

provided that in the last term, $j \neq 2i - 3$. This problem can only occur for odd j. We don't care about $G_1(R_0)$ since $W'(0) = 0$, and for $j = 3$ we get $i = 3$; so neglecting this integral in $G_3(R_0)$ leads to an error in $\delta(R_0)$ of $O(\gamma R_0^5 \ln R_0)$. For $j > 3$, these logarithmic terms are even smaller.

Although the logarithmic terms are neglectable, we can see that for odd j, some derivative of $G_j(R_0)$ will have the form K/R_0. Thus $G_j(R_0)$ is analytic at $R_0 = 0$ only for even j. This is why we have not attempted to expand $G_j(R_0)$ as a power series in R_0.

Next we can apply the result (B10) from Appendix B to eliminate the R_0^{j+3} terms for all even j. We are not concerned with $j = 1$, and the neglect of

the equivalent $j = 3$ terms leads to an error of only $O(\gamma R_0^5)$. The higher odd j terms are even smaller. Thus, keeping only terms of $O(\gamma R_0^4)$ we get:

$$G_j(R_0) = \frac{1}{j+3} - \frac{1}{2(j+1)} R_0^2 + \frac{1}{8(1-j)} R_0^4 + O(\gamma R_0^6 \ln R_0) \qquad (70)$$

which is good for all $j \neq 1$.

We are now ready to sum out (65). To do this we observe from (62) and (58) that

$$\int_0^1 R^2 W(R) dR = \sum_{j=0}^{\infty} \frac{1}{j!(j+3)} W^{(j)}(0) = \frac{1}{3} \qquad (71)$$

$$\int_0^1 W(R) dR = \sum_{j=0}^{\infty} \frac{1}{j!(j+1)} W^{(j)}(0) \qquad (72)$$

$$\int_0^1 \frac{W(R) - W(0)}{R^2} dR = \sum_{j=2}^{\infty} \frac{1}{j!(j-1)} W^{(j)}(0) \qquad (73)$$

Putting (70-3) into (65) results in:

$$\delta(R_0) = \delta_0' R_0 - \gamma h R_0^3 + O(\gamma R_0^5 \ln R_0) \qquad (74)$$

where

$$\delta_0' = 3\gamma \int_0^1 W(R) dR \qquad (75)$$

$$h = \frac{3}{4} \left[W(0) + \int_0^1 \frac{W(0) - W(R)}{R^2} dR \right] \qquad (76)$$

This is it. The deflection is essentially a simple cubic function of R_0, whose coefficients are obtainable from any given solar model by straightforward integrations. The fractional error is of $O(R_0^4 \ln R_0)$, which for $R_0 = 0.1$ is only 2.3×10^{-4}, and even for $R_0 = 0.2$ is only 2.6×10^{-3}. Although the integral in (76) may look improper, it does not give any trouble, since at $R = 0$ the integrand is just $\frac{1}{2} W''(0)$. A method for obtaining this is discussed in Appendix A.

7. Polynomial Models

In conjunction with the first order theory just developed, it is instructive to consider some very simple mass models of the sun. These are easy to calculate and understand, and exhibit many features of the (presumably) more realistic models of Appendix A. Having worked out the first order theory in terms of the density expansion (62) suggests looking at polynomial density models, i.e., those obtainable by truncating (62) after a few terms.

Two conditions will be imposed on these models. First, we must adjust the $W^{(j)}(0)$ so that $M(1) = 1$, in order that our model has the right mass. Also, from the argument in the last section, we will again assume a vanishing central density gradient; that is, $W'(0) = 0$.

The simplest model obeying these constraints is the uniform density model, for which $W_0(R) = 1$. The deflection is easy for this model. From either (61) or (74-6):

$$\delta_0(R_0) = \frac{2\gamma}{R_0} (1 - s^3 \sigma_1) = 3\gamma R_0 - \frac{3}{4} \gamma R_0^3 + O(\gamma R_0^5) \qquad (77)$$

It is not possible to construct a distinct linear model with $W'(0) = 0$, so we will go on to a quadratic model, $W_2(R)$. Here we are free to specify one more parameter; so, to add a touch of realism, we will require $W_2(1) = 0$. It is then readily shown from (58) that

$$W_2(R) = \frac{5}{2}(1 - R^2) \tag{78}$$

Again using either (61) or (74-6) we get:

$$\delta_2(R_0) = \frac{2\gamma}{R_0}(1 - s^5\sigma_1) = 5\gamma R_0 - \frac{15}{4}\gamma R_0^3 + 0(\gamma R_0^5) \tag{79}$$

A cubic model is about as far as we need to go, since we can get a fair resemblance to the published models by adding an R^3 term and requiring both $W_3(1) = W_3'(1) = 0$. Applying this to (58) we get this time:

$$W_3(R) = 5(1 - 3R^2 + 2R^3) \tag{80}$$

and again using (61) or (74-6):

$$\delta_3(R_0) = \frac{2\gamma}{R_0}\left[1 - (1 - \frac{13}{4}R_0^2 + \frac{33}{8}R_0^4)s\sigma_1 + \frac{15}{16}R_0^6\ln\left(\frac{1+s\sigma_1}{1-s\sigma_1}\right)\right]$$

$$= \frac{15}{2}\gamma R_0 - \frac{45}{4}\gamma R_0^3 + 0(\gamma R_0^5\ln R_0) \tag{81}$$

The density models and their associated mass functions are plotted in Figure II-1, and the corresponding deflections are shown in Figure II-2. For $R_0 \geq 1$, the common result is $\delta = \delta_e = \frac{2\gamma}{R_0}$ from (49). The values of δ_0'/γ and h are collected in Table II-1 below.
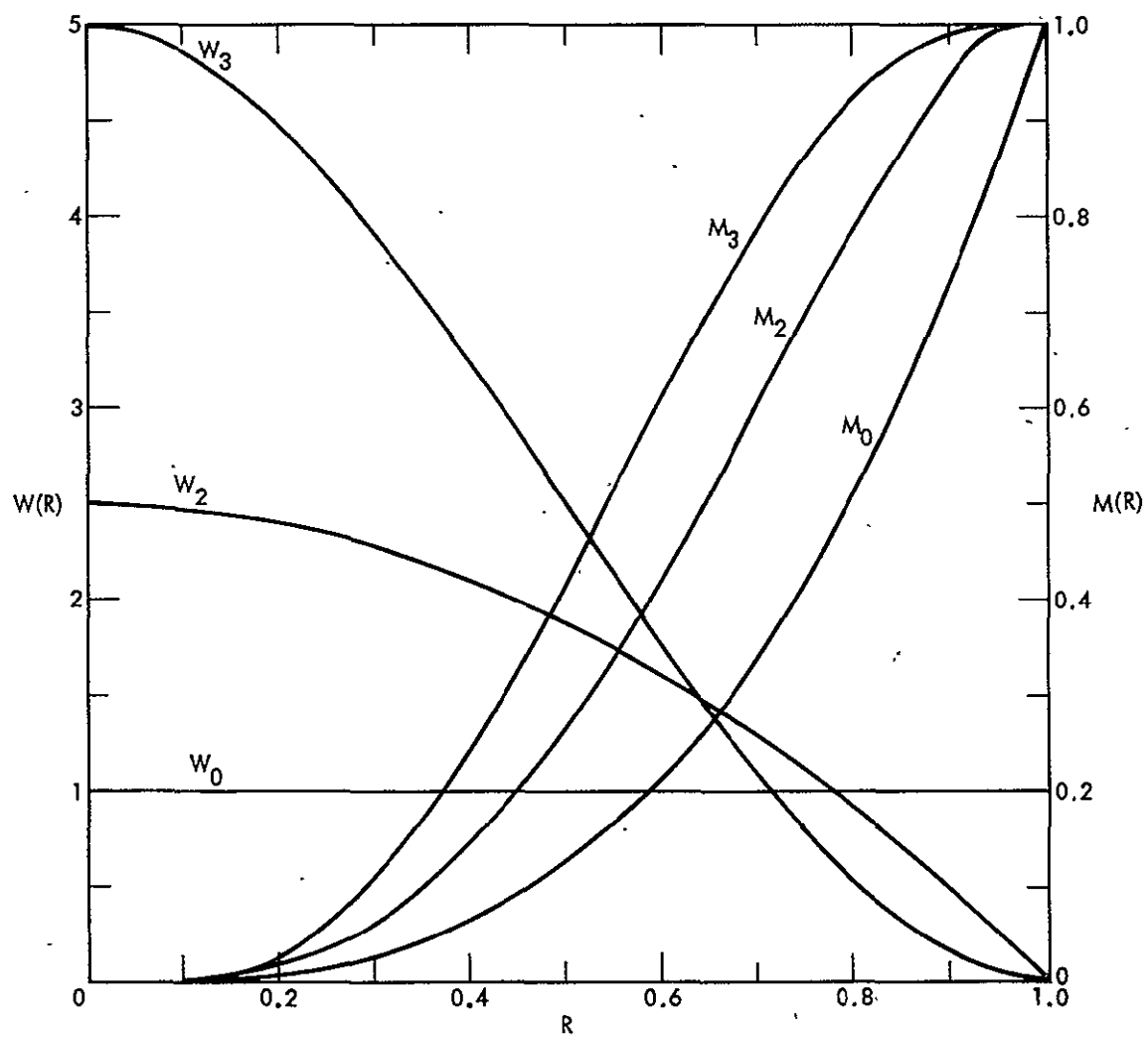
Figure II-1. Polynomial Density and Mass Models
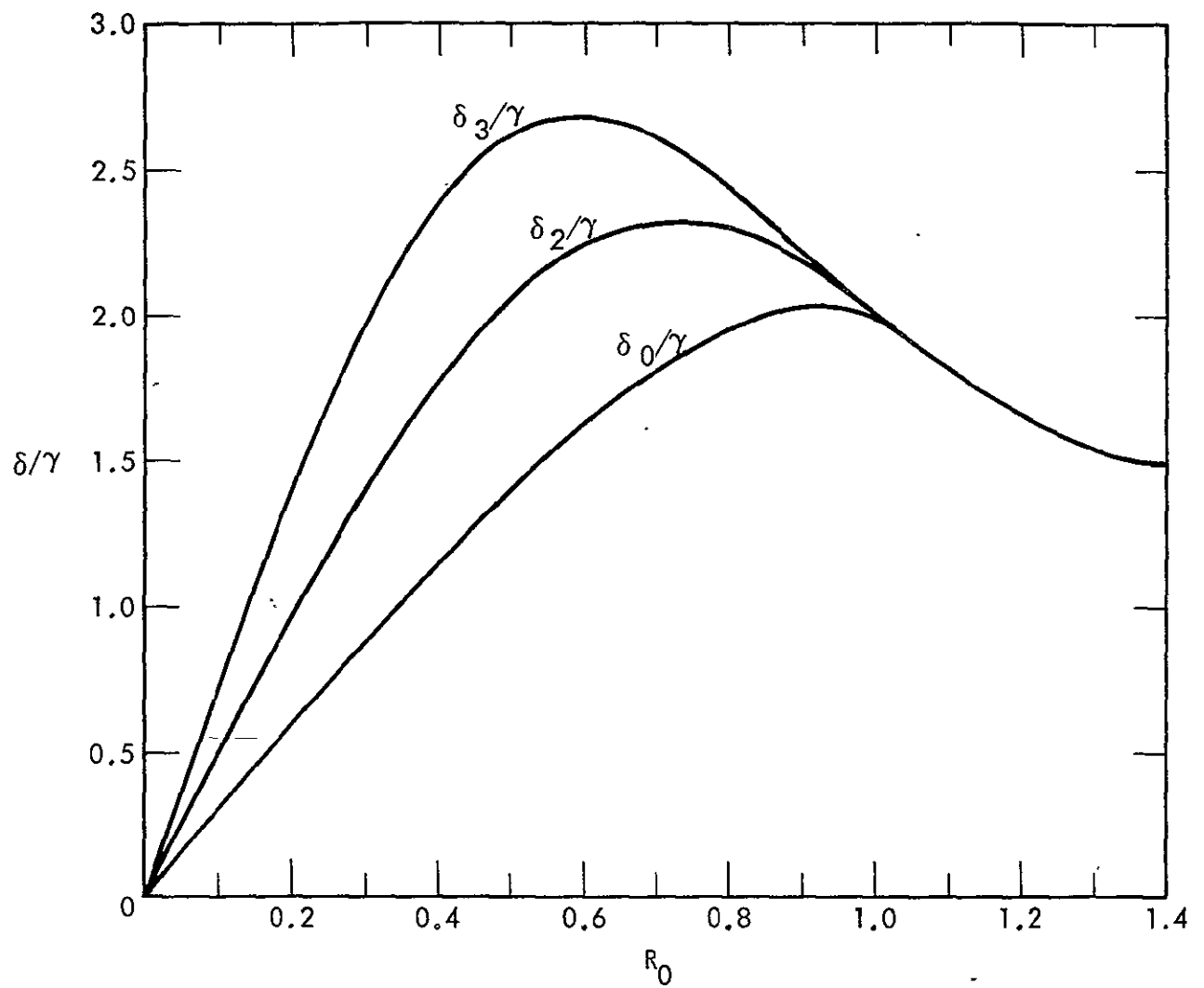
Figure II-2. Deflection vs Perihelion for Polynomial Models

It is also interesting to find the interior values of Λ and Φ for these models. From (34) we have:

$$\Lambda = \frac{\gamma}{R} M(R) + O(\gamma^2) \qquad (82)$$

Since $\Lambda'(R) < 0$ when $R > 1$, and $\Lambda(0) = 0$, we must have a maximum in the interior. This is readily calculated from (58) and (82) for each model. The values of Λ_{max}/γ and $R(\Lambda_{max})$ are included in Table II-1.

Now let's look at Φ. To first order, and neglecting the pressure, we have from (34, 5):

$$\Phi'(R) = \frac{\gamma}{R^2} M(R) \geq 0 \qquad (83)$$

Since we have chosen the usual reference value $\Phi(\infty) = 0$, $\Phi(R) \leq 0$, and the most negative value is at $R = 0$. To find this value, we can integrate (83) starting from the point

$$\Phi(1) = \frac{1}{2} \ln(1 - 2\gamma) = -\gamma + O(\gamma^2) \qquad (84)$$

which we know from (41). Thus:

$$\Phi_0 = \Phi(0) = \Phi(1) - \gamma \int_0^1 \frac{M(R)}{R^2} \, dR \qquad (85)$$

This is readily worked out for each polynomial model; and the results are included in Table II-1. It is clear that, at least for these polynomial models, our assumption that Λ and Φ are of $O(\gamma)$ is well borne out.

Table II-1. Properties of Polynomial Density Models

Model	$W(0)$	$-W''(0)$	δ_{mag}/γ	$R(\delta_{max})$	δ'_0/γ	h	Λ_{max}	$R(\Lambda_{max})$	$-\Phi_0/\gamma$	$\alpha=3\gamma h/\delta'_0$
Uniform	1	0	2.044	0.9306	3	3/4	1	1	3/2	3/4
Quadratic	5/2	5	2.331	0.7299	5	15/4	25/24	$\sqrt{5/6}$	15/8	9/4
Cubic	5	30	2.682	0.5954	15/2	45/4	1.1543	0.7764	9/4	9/2

How good are these models? Comparing Figure II-1 to Figure A1 shows that the cubic model has roughly the right shape, but the central density is far too low, and $M(R)$ climbs too slowly. A higher order model could surely give better agreement; but it is contrary to our intent of getting the main focal features from a simple model.

A better way is to merely assume that the sun is smaller than its apparent optical size. This is prompted by observing that the models of Appendix A all show $M(0.5) \sim 0.95$. All we need to do is choose a model radius $r'_s = r_s/2$, say. This increases the central density by a factor of 8, which is roughly what is needed for the cubic model. All the formulas of this section can then be used as is by merely referencing γ, R_0, and R to r'_s instead of r_s. This is why δ/γ instead of δ is used in most of the tables and graphs.

8. Solar Focusing

Consider an initially parallel bundle of rays, incident on the sun, from some remote interstellar source. One ray will pass undeflected through the center of the sun. We will call its path the optic axis, or just the axis. From the symmetry we have assumed, all the other rays will be deflected, and will cross the axis somewhere.

Suppose some given ray passes at some perihelion r_0, and later crosses the axis at radius $\ell = \ell(r_0)$. We will call ℓ the "focal length" corresponding to r_0. Consider the figure at right. Let the focus have Schwartzchild coordinates $(\ell, \phi(\ell))$. Since ϕ is measured forward from r_0, we can see that from (23)

$$\phi(\ell) = \pi - (\lambda_\infty + \delta_\infty) = \frac{\pi}{2} - \delta_\infty \qquad (86)$$
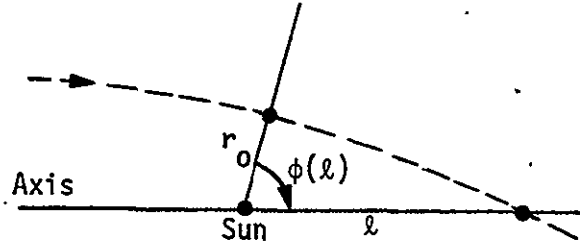
where δ_∞ is the total deflection from the source to perihelion. We can also write this as:

$$\lambda(\ell) = \phi(\ell) - \delta(\ell) = \frac{\pi}{2} - 2\delta_\infty + O(\gamma^3) \qquad (87)$$

since δ_e is constant to $O(\gamma^3)$, provided $\ell > 107$ Gm (see Section 5). We will show that this always holds.

To find ℓ we have from (15) and (45):

$$c\sigma(\ell) = \frac{r_0}{\ell} \triangleq \frac{R_0}{L} = [1 + \Phi_0 + O(\gamma^2)]c\lambda(\ell) \qquad (88)$$

provided $L > 22$, which is within the above assumptions ($L > 154$). With (87) this is:

$$\frac{R_0}{L} = 2\delta_\infty[1 + \Phi_0 + O(\gamma^2)] \qquad (89)$$

$$L = \frac{R_0}{2\delta_\infty}[1 - \Phi_0 + O(\gamma^2)] \qquad (90)$$

To get an idea of what this looks like, we can obtain δ_∞ from the simple models of the last section. Dropping the Φ_0 term, and applying (90) to the curves of Figure II-2 results in Figure II-3. For very small R_0, it is convenient to use (74):

$$L = \frac{1}{2\delta_0'}\left(1 + \frac{\gamma h R_0^2}{\delta_0'}\right) \qquad (91)$$

From (91) and Figure II-3 we can see that

$$L(R_0) \geq L(0) = (2\delta_0')^{-1} \qquad (92)$$

There is clearly no difficulty with $L > 154$ for any of our polynomial models. This is also true for the models of Appendix A, but might not hold for "unconventional" models. Observe that every point on the axis, with $R > L(0)$, is the focus for some incoming cylindrical shell of radiation. Moreover, we can see a strong pile up of rays just beyond $L(0)$. We will call this point the "central focus", and look carefully at this region in the next two sections. The optical analogy is a lens with a dreadful spherical aberration.
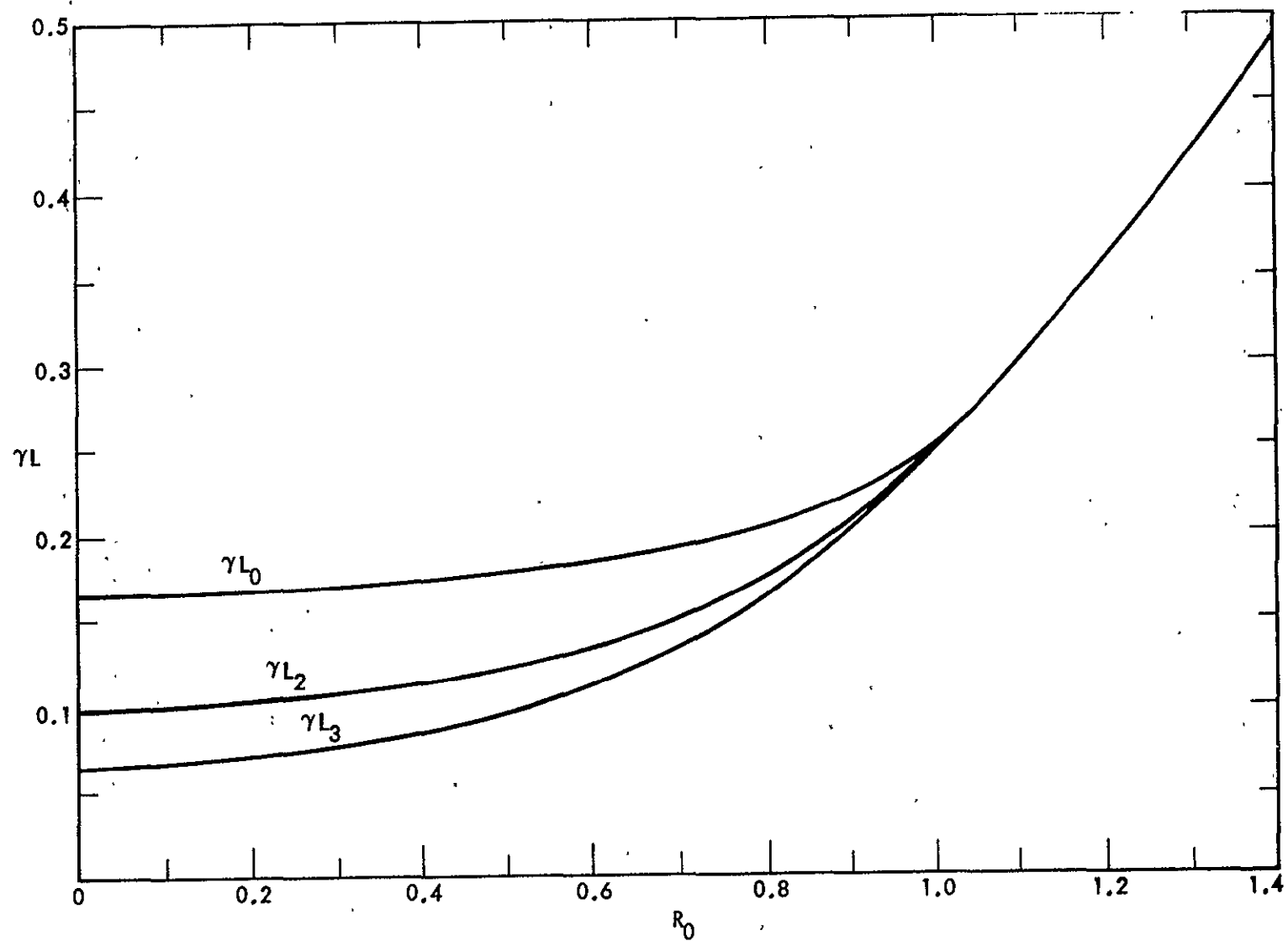
Figure II-3. Focal Length vs Perihelion for Polynomial Models

From (92) we can see that $L(0)$ is a strong function of the solar model. Here we might get an initial estimate by assuming the cubic model with $r_s' = r_s/2$. Then from Table II-1:

$$\ell_0 = r_s' L(0) = \frac{r_s}{4\delta_0'} = \frac{r_s \gamma'}{8\gamma\delta_0} = \frac{(6.9598 \times 10^8)(2)}{(8)(2 \cdot 12167 \times 10^{-6})(15)} = 5.467 \text{ Tm}$$

Between Neptune and Pluto.

This is a good point to comment on the physical significance of our solar parameters δ_0' and h. Consider (91). The prime focus $L(0)$ depends on δ_0', and nothing else. That is, if a space mission actually locates some solar focus, we will immediately have a precise determination of the density integral (75). As for h, it is evident that if $h = 0$, all rays would pass through $L(0)$. Thus h is a measure of the spherical aberration of the sun. Note that $h \neq 0$ for any of our polynomial models. Indeed, from (76), $h = 0$ would require that $W(0)$ is not the maximum of $W(R)$. As such a sun would be gravitationally unstable, we can anticipate considerable spherical aberration in the actual sun.

9. The Caustic

Consider a ray passing at some perihelion R_0, and which later reaches a radius R somewhere near $L(0)$. Let Δ be the angle off the axis which this point subtends at the center of the sun. From the figure
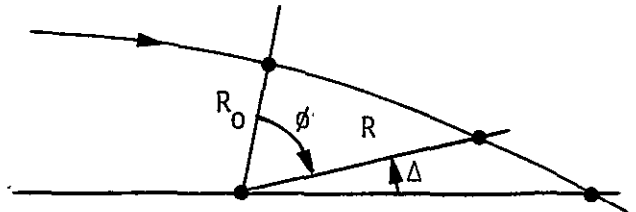
$$\Delta = \pi - \phi - (\frac{\pi}{2} + \delta_\infty) = \frac{R_0}{R} - 2\delta_\infty + O(\gamma^2) \tag{93}$$

the last from (23) and (45), after noting from (92) that $1/R = O(\gamma)$. The "distance" Y off the axis is given by:

$$Y = R\Delta + O(\gamma^2) = R_0 - 2R\delta_\infty + RO(\gamma^2) \tag{94}$$

As an example, the cubic density model is used to compute δ_∞ [(74) is adequate near $L(0)$] for different values of R_0. Each R_0 gives a straight line ray, and the family of these rays is shown plotted in Figure II-4, in the vicinity of $L(0)$. That some of the rays appear nearly vertical is due to the horizontal scale compression, γ.

The most interesting feature of Figure II-4 is that the family of rays appear to possess an envelope, after passing the axis. Optically, such a feature is called a "caustic", and we will adopt that terminology. In general, for a family of curves $F(x, y, \lambda) = 0$; an envelope obeys $\partial_\lambda F(x, y, \lambda) = 0$. These define the envelope parametrically. Applying the idea to (94) yields:

$$2R\delta_\infty'(R_0) = 1 \tag{95}$$

We will examine the caustic in the region of several hundred Gm beyond $L(0)$, where we can use (74). In this case:

$$2R(\delta_0' - 3\gamma h R_0^2) = 1 \tag{96}$$

and combining this with (94) we can show that:

$$Y = - 4\gamma h R R_0^3 \tag{97}$$
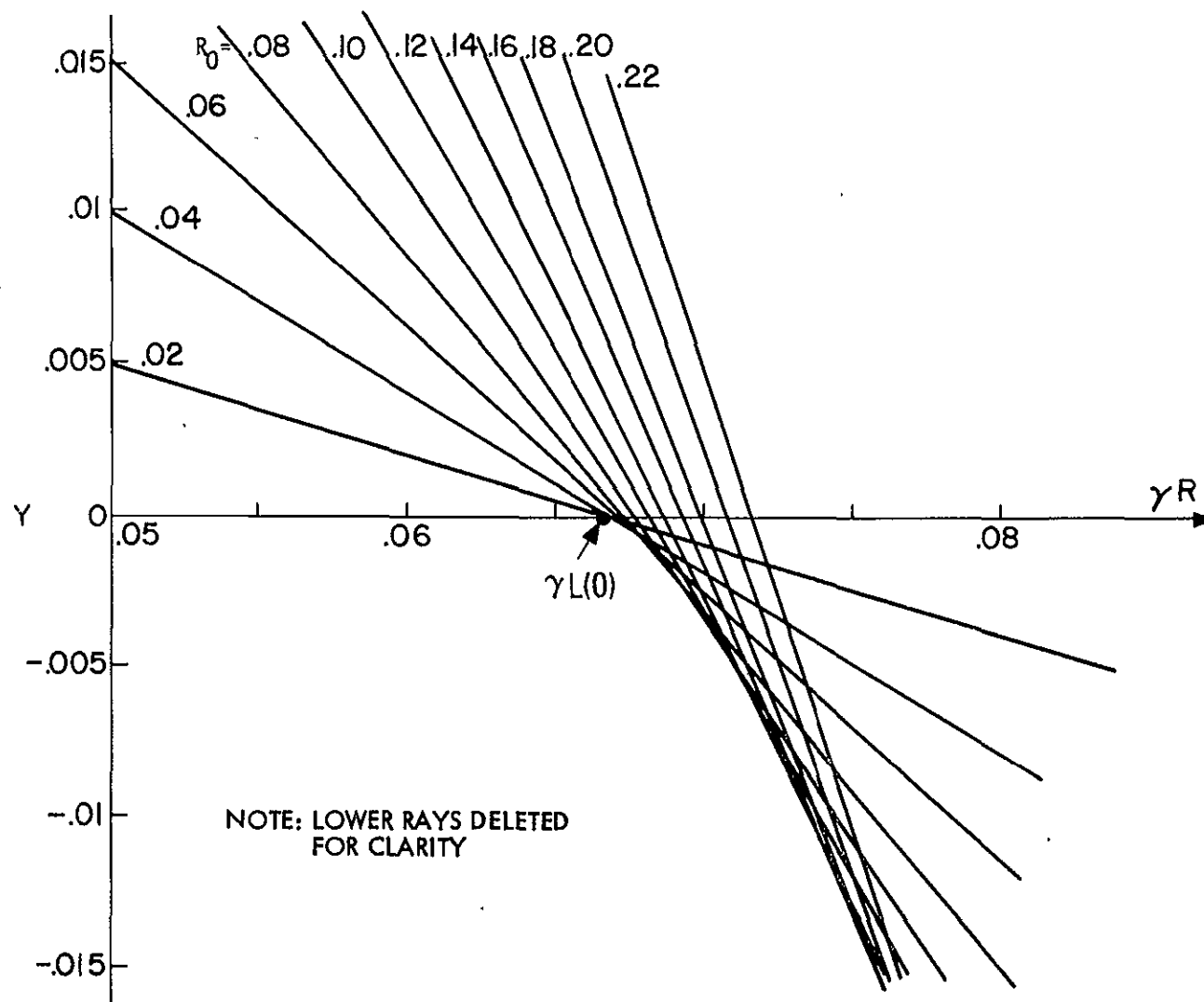
Figure II-4. Ray Trace for Cubic Density Model

We can either use these to compute the caustic parametrically from R_0, or eliminate R_0 between them and get:

$$\frac{27}{2} \gamma h R Y^2 = (2\delta_0' R - 1)^3 = \left[\frac{R}{L(0)} - 1\right]^3 \qquad (98)$$

Note that from (97), $Y'(0) = 0$, i.e. the caustic has zero slope at the central focus. A plot of the caustics for the three polynomial models is shown in Figure II-5 for $|R_0| \leq 0.14$. Beyond this point, the $O(R_0^5 \ln R_0)$ terms neglected in (74) may become important. To put these on one plot, it is helpful to begin each abscissa at $L(0)$. To this end we use (92) to reorganize (96, 7):

$$R - L(0) = \frac{1}{2\delta_0'} \left(\frac{1}{\alpha R_0^2} - 1\right)^{-1} \qquad (99)$$

$$Y = -\frac{4}{3} \delta_0' R_0 [R - L(0)] \qquad (100)$$

where

$$\alpha \equiv \frac{3\gamma h}{\delta_0'} \qquad (101)$$

While only the positive caustic has been plotted in Figure II-5, corresponding to $R_0 \leq 0$, both sides might have been shown. In fact, it is well to note that the caustic is really a surface of revolution, symmetrical about the axis. To get some idea of scale, if $r_s' = r_s/2$, then the total R range shown is 500 Gm; while the total Y scale is only 3 Mm. The vertical exaggeration is $1/\gamma' \sim 2.5 \times 10^5$.

A very important question in the next section is, given a point (R, Y), what rays pass through it? To answer this we can combine the ray equation (94) with (74) to get a cubic equation in R_0:
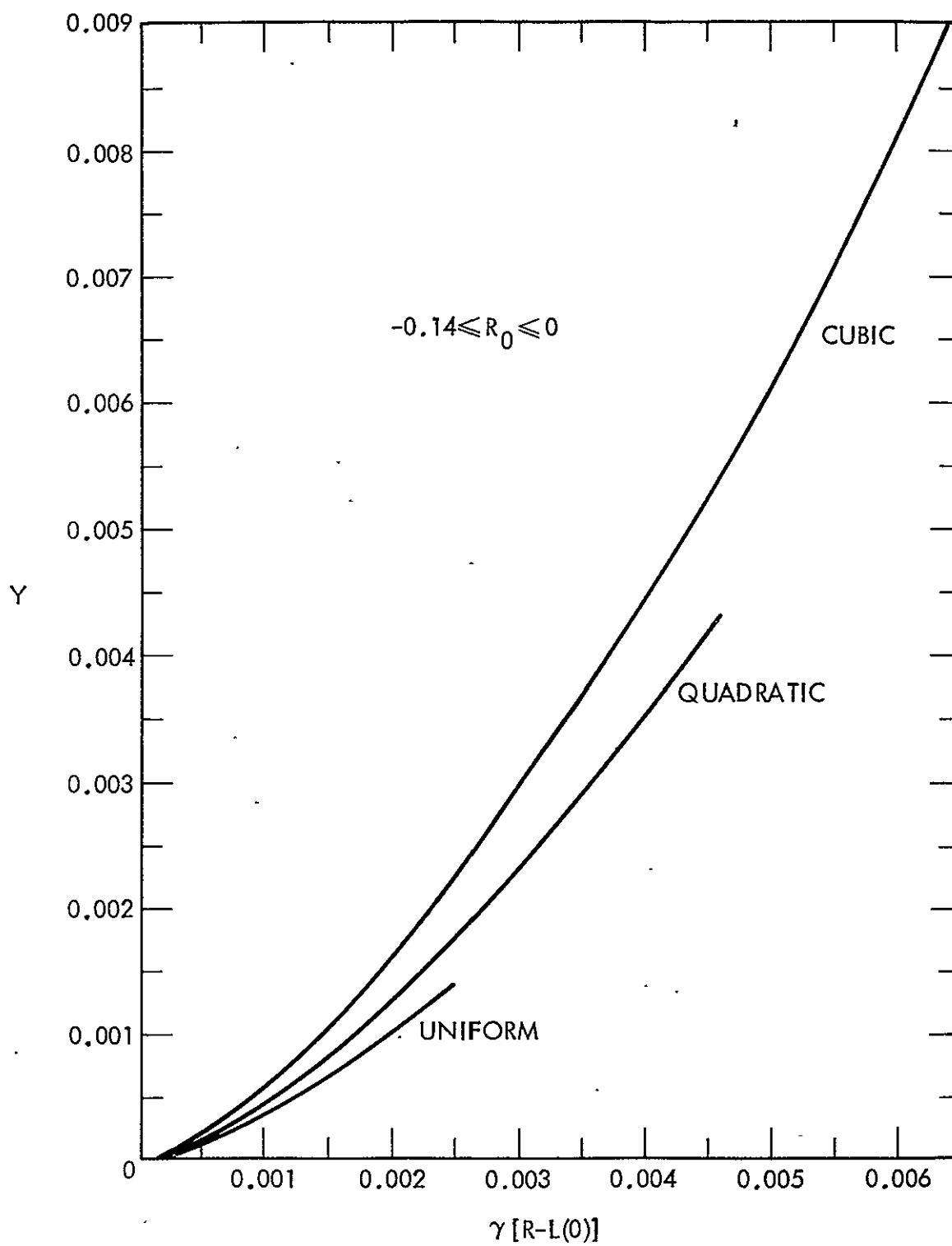
Figure II-5. Caustics for Polynomial Density Models

$$R_o^3 + \frac{1}{\gamma h} \left(\frac{1}{2R} - \delta_o' \right) R_o - \frac{\gamma}{2\gamma hR} = 0 \qquad (102)$$

One conclusion can be drawn immediately. Since a cubic equation always has at least one real solution, and as all positive and negative R_o are physically meaningful, we conclude that all points near $L(0)$ have at least one ray passing through them.

There is more to this. (102) is of the form

$$R_o^3 + aR_o + b = 0 \qquad (103)$$

so its discriminant is

$$\Delta_3 \equiv \left(\frac{a}{3} \right)^3 + \left(\frac{b}{2} \right)^2 = \left(\frac{1}{3\gamma h} \right)^3 \left(\frac{1}{2R} - \delta_o' \right)^3 + \left(\frac{\gamma}{4\gamma hR} \right)^2 \qquad (104)$$

We can see from (98) that on all points of the caustic, $\Delta_3 = 0$. Moreover, for points beyond the caustic, γ^2 is less than the caustic value; so $\Delta_3 < 0$ there, and conversely. Thus the caustic separates regions of one and three real roots, and beyond the caustic there are always three distinct rays passing through each point.

10. Gain

We are now ready to examine the degree of concentration near the focus. Suppose the incoming radiation has remote intensity I. On reaching R_o it will be concentrated by a factor $1 + O(\gamma)$, which we will ignore. Thus the total power in an incoming annulus of radius R_o and thickness dR_o is $2\pi R_o I dR_o$. When the radiation passes the sun and reaches radius R, it will

have concentrated into an area $2\pi Y dY$, apart from a factor of $c\Delta = 1' - 0(\gamma)$. Based on this rationale we define the gain at an arbitrary point as:

$$G = \sum \frac{R_o dR_o}{Y dY} \qquad (105)$$

The summation is required beyond the caustic.

From (94) the gain can be written:

$$G = \sum \left(\frac{Y}{R_o}\right)^{-1} \left(\frac{dY}{dR_o}\right)^{-1} = \sum \left[1 - 2R \frac{\delta_\infty(R_o)}{R_o}\right]^{-1} [1 - 2R\delta_\infty'(R_o)]^{-1} \qquad (106)$$

Very large gain can occur if any bracketed term in (106) vanishes. The first term vanishes on the axis at any point beyond $L(0)$, and is just the axial focus found in (90). The second term vanishes on the caustic, as is plain from (95). We can look at (106) as a sort of gravitational Guide Michelin to the solar foci.

It should be emphasized that the definition (105) is purely geometric. No recognition is made of the possible wave nature of the radiation, and intensities are combined as though from incoherent sources. This suggests that the infinities in (106) aren't physical - it might not be trustworthy closer than a few reduced wavelengths to the axial focus or the caustic. The other places to watch out for are those points beyond the caustic where the dominant contributing rays have nearly equal gains, so that interference could play a role. These considerations will be ignored in the remainder of the section.

In general, the calculation of G is rather difficult, especially beyond the caustic, where all three rays must be identified at each point. The process is greatly simplified near $L(0)$ where the approximation (74) can be used. With this, (92), and (101), we can rewrite (106) as

$$G = \sum \frac{R_0}{Y} \left[1 - \frac{R}{L(0)} \left(1 - \alpha R_0^2\right)\right]^{-1} \qquad (107)$$

Given R and Y, R_0 is the set of real solutions of (102), after which (107) is straightforward to evaluate. This process has been carried out for the cubic density model along a set of strips of constant Y, and G vs. R was plotted. A set of constant gain contours were found from this, and are shown in Figure II-6. Once again for scale, if $r_s' = r_s/2$, then a change in YR of .001 corresponds to 82.1 Gm; while Y = .003 corresponds to 1.044 Mm, for a vertical exaggeration of again about 2×10^5.
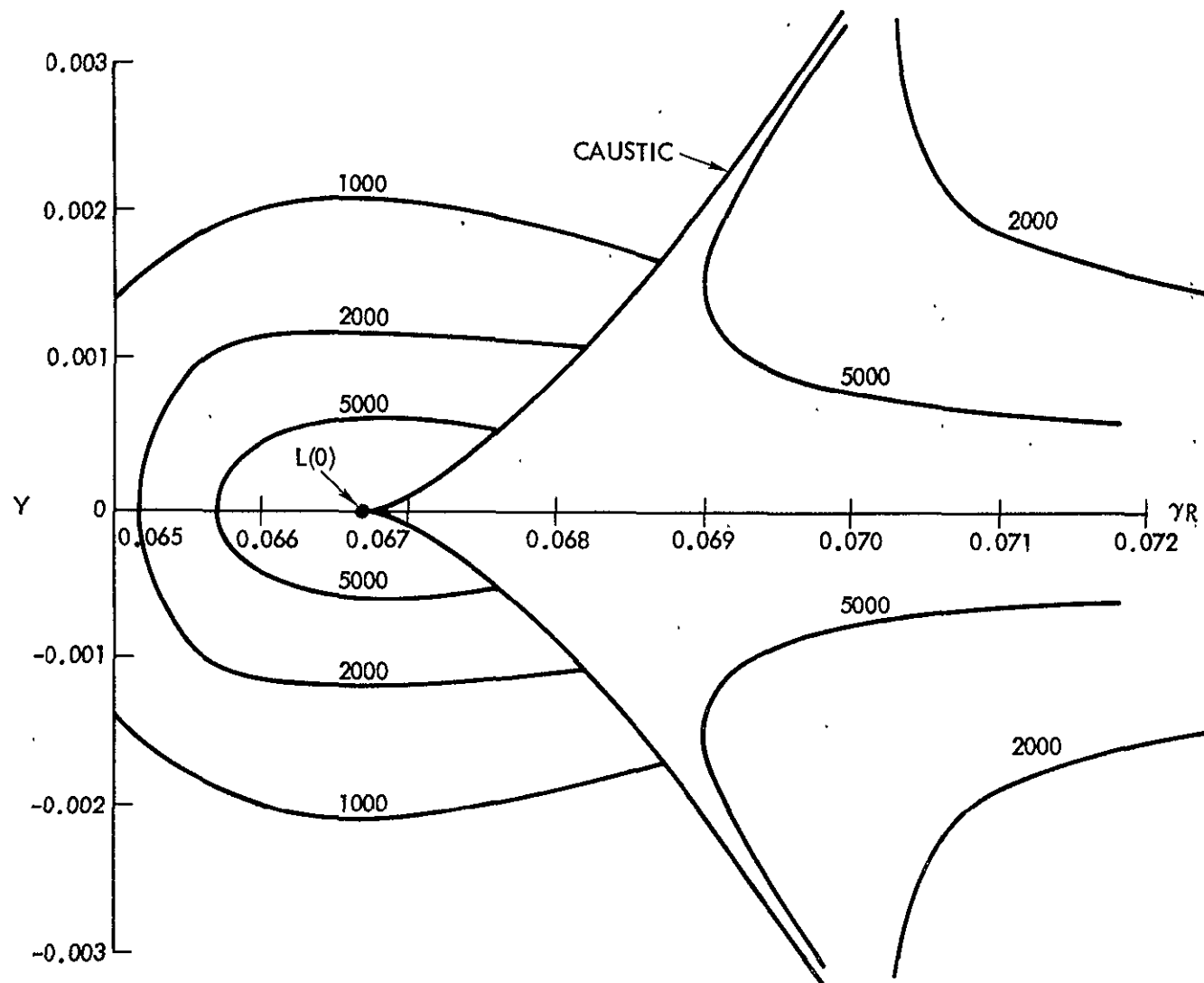
Figure II-6. Gain Contours for Cubic Density Model

CHAPTER III

TWO IMPULSE TRANSFER

1. Rationale

The physical problem of where we are going can now be regarded as settled; and we can turn to the engineering problem of how to get there. Lacking some new evidence bearing on the solar density, a substantial uncertainty in the focal radius will remain, at least until the first flight. Thus, the first mission must be prepared for a lengthy search phase in which the flight is directly along the axis.

There are two possibilities--flying outward and inward. The latter permits some propellant savings; but, as we will show, requires so much additional time as to sorely try our patience. We will only seriously consider outward trajectories, and explore their propulsion requirements.

The simplest, and least expensive, spacecraft would use only chemical rockets for their main propulsion. A mission based solely on this concept would start with a large impulse from shuttle orbit, putting the spacecraft on a long transfer orbit. This event will be called "departure". On reaching the optic axis at a predetermined "turn point", a second impulse would be fired to give the spacecraft a predetermined, purely axial, velocity. This scenario will be referred to as a "two impulse transfer". Our objective, in this chapter, is to optimize the two impulses so as to find the maximum payload for a given optic axis direction. Total mission time will also be a factor.

An actual mission, relying solely on chemical propulsion, is not necessarily limited to two impulses. Many studies have shown that extra impulses often yield significant improvements in payload for planetary transfers. Of even greater value are the possibilities for "gravity assists". These are precisely controlled encounters with planets to give useful changes of velocity. Perhaps the outstanding example of this was Mariner/Venus/Mercury, with five planetary interactions, three of which were controlled "assists". Unfortunately, time and space do not permit explorations of these possibilities here; future research will almost surely show the present conclusions to be conservative.

The optimization of two impulse transfers is an old problem, in connection with orbit changes and interplanetary trajectories. However, the present requirement of terminating on a purely axial trajectory (zero angular momentum) appears to be unique, and the literature has not been of much help. Accordingly, we will treat this as a classical problem in orbit theory, followed by a payload optimization, using standard methods.

These ground rules have been adopted for the study:

A.   The original spacecraft is a single dedicated Shuttle payload in low earth orbit. The maximum Shuttle capability is assumed, to give the maximum final payload.

B.   Presently available, Shuttle compatible propellants are assumed for both burns. The turn point propellants must be storable as well.

C.   Equipment unnecessary to the final mission phase may be jettisoned during the transfer orbit, and empty tankage may be jettisoned

even during the burns. These possibilities are discussed in Section 7.

D. The earth's orbit is assumed circular at the earth-sun mean distance, and planetary perturbations are ignored. For the present purpose of establishing performance, these approximations are not severe; and they greatly simplify the analysis. Moreover, the average performance bias due to this neglect should be negligible.

E. The optimization is over all possible impulse magnitudes and directions at both burns, all positions of the earth in its orbit at departure, and all arrival times at the turn point, consistent with the constraints of celestial mechanics.

F. The universe will be assumed newtonian, and to consist purely of the sun, the spacecraft, and the geometric notions of the earth's orbit and the optic axis. The gravitational effect of the earth is included in the departure analysis of Section 6.

G. The distance r_T of the turn point and the axial speed v_T just after the turn are regarded as constants for the optimization. A range of values will be considered in the numerical search; and the factors bearing on their choice, particularly mission time and focal point uncertainty, will get considerable attention.

2. Free Axial Flight

Perhaps the best way to get an initial feeling for the numbers is to examine the search phase of the mission. This is where we coast out the optic

axis, beyond the turn point. Closed form time vs. radius expressions exist here, and the relations between the parameters can be plotted fairly clearly.

A convenient parameter for describing these trajectories is the spacecraft total specific energy. This constant of the motion is given by:

$$E = \frac{1}{2} v^2 - \frac{\mu}{r} = \frac{1}{2} v_T^2 - \frac{\mu}{r_T} \tag{1}$$

where r is the distance from the sun, v is the speed, r_T and v_T are the values just after second burn, and[1]:

$$\mu = Gm_s = 1.327124 \times 10^{20} \; m^3/sec^2$$

A series of energy contours are plotted in Figure III-1.

In purely axial motion, $v = \dot{r}$, so (1) integrates to:

$$t - t_T = \frac{1}{\sqrt{2}} \int_{r_T}^{r} (E + \frac{\mu}{x})^{-1/2} dx \tag{2}$$

No single expression seems possible for this integral; but for the various E possibilities we get:

$$t - t_T = \begin{cases} \frac{1}{E}\sqrt{\frac{x}{2}} \; (\mu + Ex) - \mu(-2E)^{-3/2} s^{-1}(1 + \frac{2Ex}{\mu}) & x=r \quad (E<0) \\[2ex] \frac{1}{3}\sqrt{\frac{2}{\mu}} x^{3/2} & (E=0) \\[2ex] \frac{1}{E}\sqrt{\frac{x}{2}} \; (\mu + Ex) - \mu(2E)^{-3/2}\ln[\sqrt{Ex(\mu + Ex)} + Ex + \frac{1}{2} \mu] & x=r_T \quad (E>0) \end{cases} \tag{3}$$

These expressions are plotted in Figure III-2, for various E values.
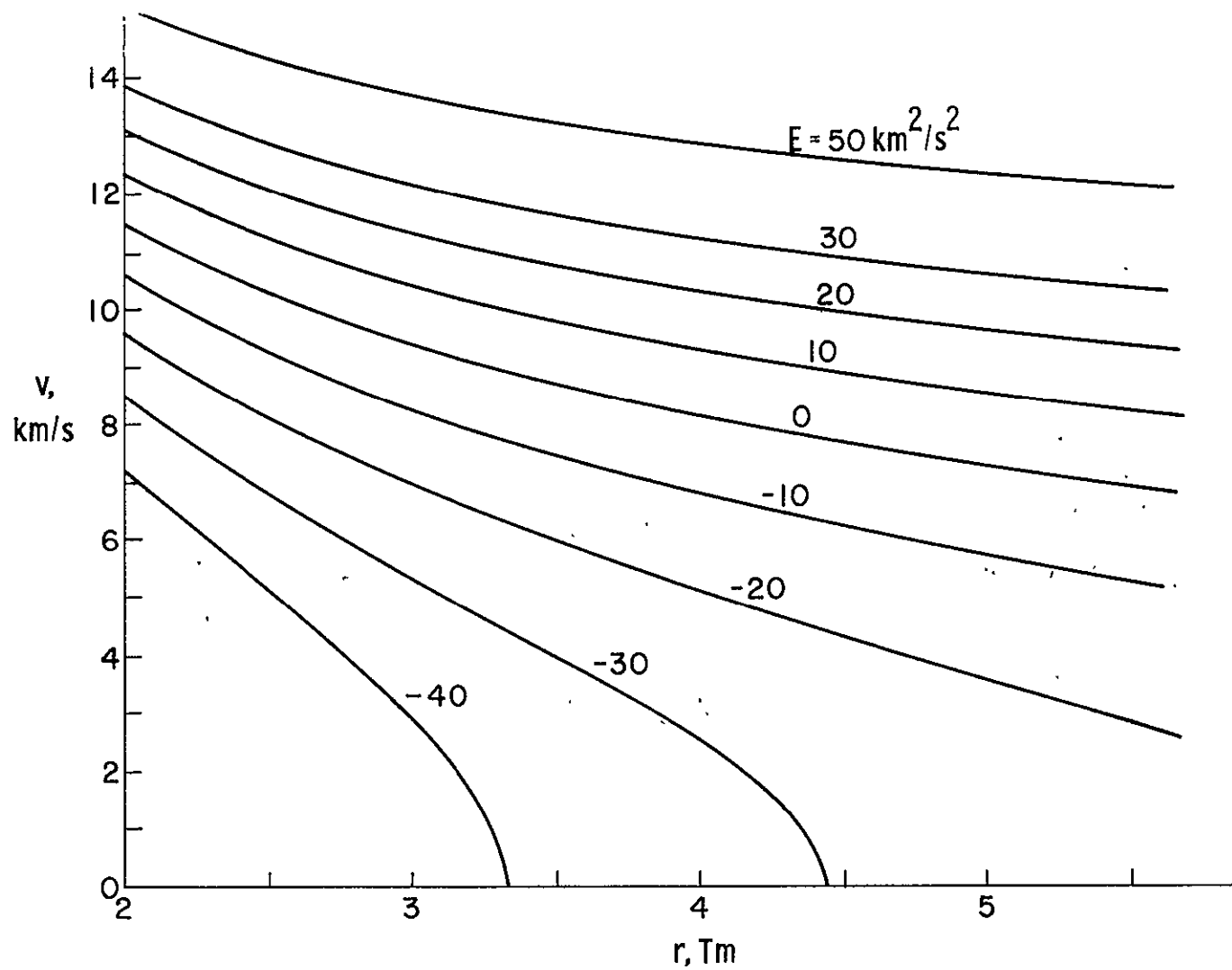
Figure III-1. Energy Equation Contours
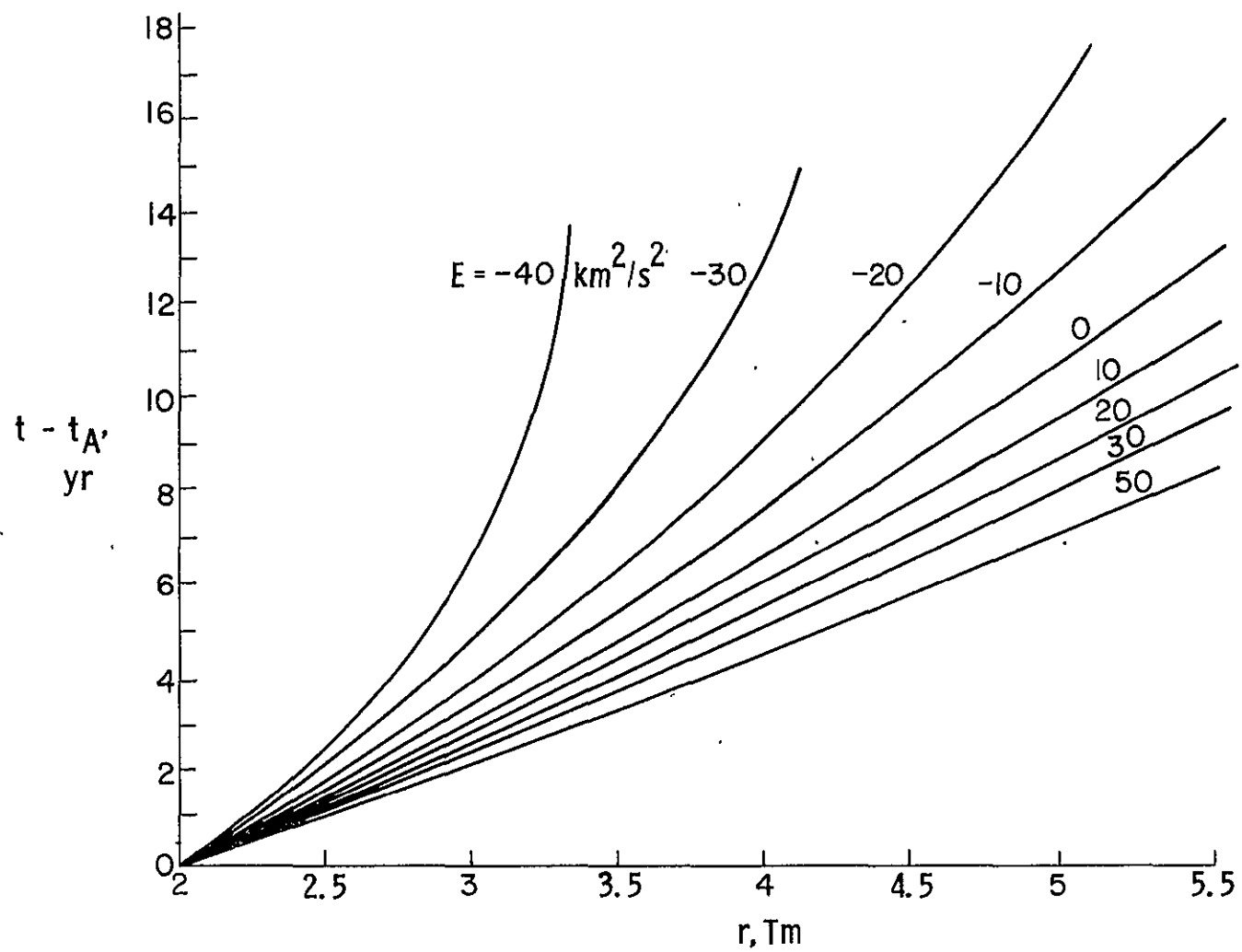
Figure III-2. Axial Flight Time

From Figure III-1, we can see that if the axis is to be explored out to 5.5 or 6 Tm, we will need $E > - 25$ km^2/sec^2, which implies $v_T > 9$ km/sec for $r_T \sim 2$ Tm. Lack of patience may set tighter limits. From Figure III-2, if we allow, say, 15 years to traverse the range $2 < r < 5.5$ Tm, then we require $E > - 9$ km^2/sec^2, for which $v_T \sim 11$ km/sec at $r_T = 2$ Tm. Further time reductions can be very expensive.

3. Definitions and Coordinate Systems

Four different coordinate systems play roles in our analysis. Each system is characterized by a set of right handed orthonormal base vectors E_α^X; where $X = G, S, N,$ or T identifies the system; and $\alpha = 1, 2,$ or 3 identifies which vector.

The fundamental system is E^S, and is based on the ecliptic. E_1^S is defined to be the direction to the vernal equinox, and E_3^S will be taken as the north ecliptic pole. A more precise definition would be required for mission analysis, but we won't need it here.

The next system, E^G, is a standard non-rotating earth system. Following the usual astrodynamic convention we take $E_1^G = E_1^S$, and E_3^G as the north polar axis. Letting[1] $\varepsilon = 23.5$ deg be the obliquity of the ecliptic, these are related by:

$$E_\alpha^S = R_{\alpha\beta}^1 (\varepsilon) E_\beta^G \tag{4}$$

Here we have adopted the nomenclature $R^\Lambda(\theta)$ to be the matrix representing a rotation through the angle θ about the axis E_Λ.

The next system is E^N, which we will call the nodal system. It is ecliptic like E^S; but it is based on the location of the earth at departure, rather than the vernal equinox. If the earth is at a longitude ϕ_e relative to E_1^S at departure, then:

$$E_\alpha^N = R_{\alpha\beta}^3(\phi_e)E_\beta^S \tag{5}$$

From this, we can characterize the earth's position as:

$$\bar{r}_e = r_e E_1^N \tag{6}$$

where[1] $r_e = 149.5979$ Gm. Similarly, the earth's velocity is[1]:

$$\bar{v}_e = v_e E_2^N \tag{7}$$

$$v_e = \sqrt{\mu/r_e} = 29.78468 \text{ km/sec.} \tag{8}$$

Finally, we must bring the optic axis into the picture. We suppose that the outward axis is at a latitude θ_a and a longitude ϕ_a, relative to E^S. (These are fixed numbers for any given stellar source.) Then a unit vector $\hat{U}^a$ along this axis has components

$$\hat{U}^a = R^3(-\phi_a)R^2(\theta_a)E_1^S = [Ac\phi_a, As\phi_a, s\theta_a]^S \tag{9}$$

$$= R^3(-\phi)R^2(\theta_a)E_1^N = [Ac\phi, As\phi, s\theta_a]^N \tag{10}$$

where

$$A \equiv c\theta_a, \qquad \phi \equiv \phi_a - \phi_e \tag{11}$$

Here we have adopted the practice of adding a superscript to identify which coordinate system we are taking components in. Also, column vectors have been written as row vectors to save space; but this should cause little confusion.

The mission phase between departure and the turn point will be referred to as the "transfer orbit". Since it must pass through the points $\bar{r}_e$ and $r_T\hat{U}^a$, we can conclude that the transfer orbit plane contains both E_1^N and $\hat{U}^a$. Thus the transfer orbit normal $\hat{U}^n$ is along $E_1^N \times \hat{U}^a$. Computing this from (10), and normalizing gives:

$$\hat{U}^n = Q^{-1}[0, -s\theta_a, As\phi]^N \tag{12}$$

$$Q \equiv \sqrt{1 - A^2c^2\phi} \tag{13}$$

This shows that the transfer orbit is inclined to the ecliptic by an angle I given by:

$$cI = E_3^N \cdot \hat{U}^n = AQ^{-1}s\phi \tag{14}$$

Unlike the usual astrodynamic convention, we will take $I < 0$ when $\hat{U}^a$ is south of the ecliptic.

This brings us finally to E^T, the transfer system. It is based on taking $E_1^T = E_1^N$ and $E_3^T = \hat{U}^n$. It obeys:

$$E_\alpha^T = R_{\alpha\beta}^1(I)E_\beta^N \tag{15}$$

It will prove useful in analyzing departure and the transfer orbit.

4. The Transfer Orbit

Following departure the heliocentric orbit will be some kind of conic section. Thus:

$$\frac{\ell}{r} = 1 + ec\theta \qquad (16)$$

where r is the distance from the sun, θ is the plane polar angle measured from perihelion, e is the eccentricity, and ℓ is the semi-latus rectum. There is one geometric constraint on ℓ and e. (16) must be satisfied at departure (r_e, θ_D), and at the turn point (r_T, θ_T); while $\theta_T - \theta_D$ must satisfy:

$$c(\theta_T - \theta_D) = \hat{U}^a \cdot E_1^N = Ac\phi \qquad (17)$$

We may also write this as:

$$s(\theta_T - \theta_D) = \pm Q \qquad (18)$$

The ambiguity in (18) can be resolved by examining the quadrant of ϕ.

Next, we consider the energy equation:

$$E = \frac{1}{2} v^2 - \frac{\mu}{r} = \frac{1}{2} v_D^2 - v_e^2 = \frac{1}{2} v_a^2 - \frac{\mu}{r_T} = - \frac{\mu}{2a} = \frac{\mu}{2\ell} (e^2 - 1) \qquad (19)$$

Here E is the total specific energy of the transfer orbit, and v_D and v_a are the heliocentric speeds at departure and arrival at the turn point respectively, while a is the semi-major axis. We will find it useful to resolve $\bar{v}_D$ and $\bar{v}_a$ into radial and normal components, defined by:

$$\overline{v}_D = v_r E_1^T + v_h E_2^T = [v_r, v_h, 0]^T \qquad (20)$$

$$v_{\parallel} = \overline{v}_a \cdot \hat{U}^a; \qquad v_a^2 = v_{\parallel}^2 + v_{\perp}^2 \qquad (21)$$

These allow concise expressions for the orbital angular momentum:

$$\overline{h} = \overline{r} \times \overline{v} = h E_3^T \qquad (22)$$

$$h = r_e v_h = r_T v_{\perp} = \sqrt{\mu \ell} \qquad (23)$$

· The last is a well known result in orbit theory.

Another relation amongst these variables comes from differentiating (16):

$$\frac{\ell}{r^2} dr = es\theta d\theta \qquad (24)$$

Now, since at departure $v_r = \dot{r}$ and $v_h = r\dot{\theta}$, we can write this as:

$$\frac{\ell}{e} = \frac{r^2 s\theta d\theta}{dr} = \frac{v_h}{v_r} r_e s\theta_D = \frac{v_{\perp}}{v_{\parallel}} r_T s\theta_T \qquad (25)$$

Before proceeding to organize all this, we can reduce the symbol clutter somewhat by non-dimensionalizing. We define:

$$\alpha = \frac{r_e}{r_T}; \quad \beta = \frac{v_r}{v_e}; \quad \gamma = \frac{v_h}{v_e}; \quad \sigma = \frac{v_{\parallel}}{v_e} \qquad (26)$$

Then, from (8) and (23):

$$\ell = r_e \gamma^2; \quad v_{\perp} = \alpha \gamma v_e \qquad (27)$$

For size, we note that if $r_T = 2$ Tm, then $\alpha = .074799$.

Using all these, and (21), the energy equation (19) becomes:

$$\beta^2 + \gamma^2 - 2 = \sigma^2 + \alpha^2\gamma^2 - 2\alpha = - \frac{r_e}{a} = \frac{e^2-1}{\gamma^2} \qquad (28)$$

The conic section relation (16) turns into:

$$ec\theta_D = \gamma^2 - 1; \qquad ec\theta_T = \alpha\gamma^2 - 1 \qquad (29)$$

and the slope relation (25) becomes:

$$es\theta_D = \gamma\beta; \qquad es\theta_T = \gamma\sigma \qquad (30)$$

These relations are not all independent, as (28) is derivable from (29) and (30). However, an independent relation involving only β, γ, and σ is available by combining (17), (29), and (30):
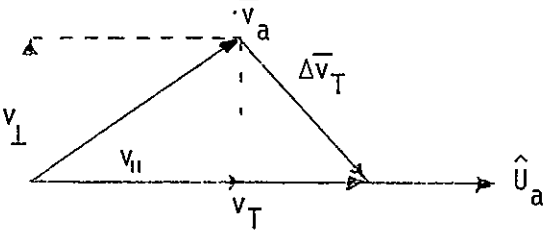
$$Ae^2c\phi = A[(\gamma^2 - 1)^2 + \gamma^2\beta^2]c\phi = (\gamma^2 - 1)(\alpha\gamma^2 - 1) + \gamma^2\beta\sigma \qquad (31)$$

We will base our optimization on (28) and (31). Most of the other relations will be useful in sorting out the geometry.

5. The Turn Point Impulse

As part of the payload optimization, we will need an expression for the velocity increment Δv_T that we must apply at the turn point in order to enter the search phase of the mission. Since we are regarding the final axial speed v_T as given, Δv_T depends only on the arrival velocity $\overline{v}_a$.

From the figure:

$$(\Delta v_T)^2 = (v_T - v_{\shortparallel})^2 + v_{\perp}^2 \tag{32}$$

As before, we will normalize this to v_e. From (26) and (27):

$$v^2 \equiv \left(\frac{\Delta v_T}{v_e}\right)^2 = (\tau - \sigma)^2 + \alpha^2\gamma^2 \tag{33}$$

$$\tau \equiv v_T/v_e \tag{34}$$

This is the form we will need.

6. Departure

As in the last section, we need an expression for the velocity increment Δv_D. This is a much harder problem, since Δv_D must take us from low earth orbit to the transfer orbit injection condition $\bar{v}_D$. Let's begin by describing the Shuttle orbit. The most efficient Shuttle injection is due east out of Kennedy Spaceflight Center (latitude $\theta_K = 28.5^0$), when it can put its maximum payload of 65,000 lb = 29,484 kg into a circular orbit of 100 nautical mile altitude [84]. Assuming [1] an earth mean radius of 6.37103 Mm, the orbit radius is:

$$r_o = 6.37103 + (100)(1.85325 \times 10^{-3}) = 6.55636 \text{ Mm}$$

and the orbit speed is [1]:

$$v_0 = \sqrt{\mu_e/r_0} = \sqrt{3.98603 \times 10^5/6556.36} = 7.797204 \text{ km/sec.}$$

Of primary concern to us is the Shuttle orbit normal, $\hat{U}^0$. If the nodal longitude relative to E_1^G is Ω, then:

$$\hat{U}^0 = R^3(-\Omega)R^1(-\theta_K)E_3^G = [s\theta_K s\Omega, \; -s\Omega_K c\Omega, \; c\theta_K]^G$$

$$= [s\theta_K s\Omega, \; s\epsilon c\theta_K - c\epsilon s\theta_K c\Omega, \; c\epsilon c\theta_K + s\epsilon s\theta_K c\Omega]^S \tag{35}$$

Here, ϵ and θ_K are fixed; but we are free to choose any Ω by an appropriate choice of the time of day of launch. Geometrically, the achievable directions $\hat{U}^0$ form a cone of half angle θ_K about E_3^G.

Consider now the desired departure velocity $\overline{v}_D$. From (20), (26), and (15)

$$\overline{v}_D = v_e[\beta, \; \gamma, \; 0]^T = v_e[\beta, \; \gamma cI, \; \gamma sI]^N \tag{36}$$

Actually, on escaping from the earth, the spacecraft already possesses a velocity $\overline{v}_e$. Thus the objective of the impulse Δv_D is to produce a hyperbolic excess velocity relative to the earth of:

$$\overline{v}_H = \overline{v}_D - \overline{v}_e = v_e[\beta, \; \gamma cI - 1, \; \gamma sI]^N \tag{37}$$

At this point it is convenient to introduce a new parametrization of the ecliptic components of $\overline{v}_H$. We define q and ϕ_0 such that:

$$q^2 = \beta^2 + (\gamma cI - 1)^2; \quad \beta = qs\phi_0; \quad \gamma cI - 1 = qc\phi_0 \tag{38}$$

with these we have:

$$\overline{v}_H = v_e[qs\phi_0, \; qc\phi_0, \; \gamma sI]^N = v_e[qs\phi_1, \; qc\phi_1, \; \gamma sI]^S \tag{39}$$

where

$$\phi_1 \equiv \phi_0 - \phi_e \tag{40}$$

We will normalize this also:

$$\rho^2 \equiv \frac{1}{v_e^2} |\overline{v}_H|^2 = q^2 + \gamma^2 s^2 I = \beta^2 + \gamma^2 - 2\gamma cI + 1 \tag{41}$$

If the spacecraft did not have to climb out of the earth's gravity well, we could minimize Δv_D by arranging $\overline{v}_0$ to be parallel to $\overline{v}_H$, in which case, $\Delta v_D = v_H - v_0$. More realistically, Δv_D is minimized if we can make $\overline{\Delta v_D}$ parallel to $\overline{v}_0$, and still achieve the required $\overline{v}_H$. If this is possible, $\overline{v}_H$ will lie in the original Shuttle orbit plane. We conclude that the optimum condition, $\overline{\Delta v_D}$ parallel to $\overline{v}_0$, is achievable if we can make $\hat{U}^0$ orthogonal to $\overline{v}_H$.

To explore this, we form the function

$$F = \frac{1}{v_e} \overline{v}_H \cdot \hat{U}^0 \tag{42}$$

From (35) and (39), after some rearrangement:

$$F = qs\phi_1 s\theta_K s\Omega + (\gamma s\epsilon sI - qc\epsilon c\phi_1)s\theta_K c\Omega + (\gamma c\epsilon sI + qs\epsilon c\phi_1)c\theta_K \tag{43}$$

This looks formidable; but we can make progress if we write the orthogonality condition $F = 0$ as:

$$Bs\Omega + Cc\Omega + D = 0 \tag{44}$$

Letting

$$c\Lambda = \frac{B}{\sqrt{B^2 + C^2}}; \quad s\Lambda = \frac{C}{\sqrt{B^2 + C^2}} \tag{45}$$

we find:

$$s(\Omega + \Lambda) = -\frac{D}{\sqrt{B^2 + C^2}} \tag{46}$$

Thus, (44) possesses real solutions provided

$$B^2 + C^2 \geq D^2 \tag{47}$$

Replacing the original expressions from (43), and after some algebra:

$$\rho s\theta_K \geq |\gamma c\epsilon s I + q s\epsilon c\phi_1| \equiv G(\phi_1) \tag{48}$$

The geometric interpretation of all this is that the set of all directions orthogonal to $\overline{V}_H$ is a plane. Optimality is achievable if this plane intersects the cone of possible $\hat{U}^0$ directions; which in turn requires that (48) be satisfied.

Since $\theta_K > \epsilon$, and from (41), $\rho \geq q$, we can see that for sufficiently small $|I|$, (48) is satisfied for all ϕ_1 (or ϕ_e). This situation is illustrated at right for $I > 0$. However, for a somewhat larger I, (48) is violated for a region around $\phi_1 = 0$. The situation for $I < 0$ is

III-16
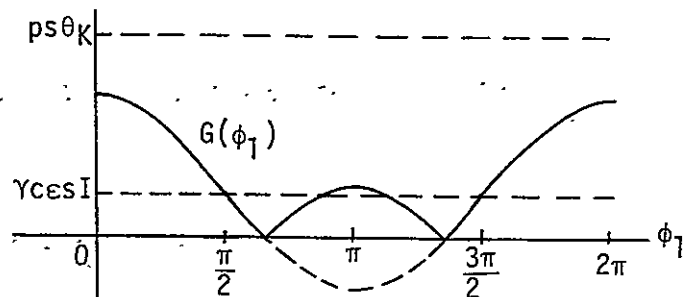
the same except that now the hump around $\phi_1 = \pi$ is larger, and the violation region shifts accordingly. Finally, if $|I|$ is big enough, $G(\phi_1)$ lies entirely above $\rho s \theta_K$, and (48) cannot be satisfied for any $\dot{\phi}_1$.

To sort this all out we define

$$c\phi_2 \equiv \frac{\rho s \theta_K - \gamma c \varepsilon s |I|}{q s \varepsilon} \tag{49}$$

when $c\phi_2 \geq 1$, (48) is satisfied for all ϕ_1. When $c\phi_2 < -1$, (48) fails for all ϕ_1. Finally, when $-1 \leq c\phi_2 < 1$, we choose the solution ϕ_2 such that $0 < \phi_2 \leq \pi$. Then (48) fails for $|\phi_1| < \phi_2$ if $I > 0$, and for $|\pi - \phi_1| < \phi_2$ if $I < 0$. When we later find optimum values for β, γ, and ϕ for a given θ_a, we will use these relations to find the range of ϕ_a for which (48) can be satisfied.

The value of Δv_D can now be found. Assuming optimality in the above sense, we can write an energy equation relative to the earth:

$$E_e \equiv \frac{1}{2} (v_0 + \Delta v_D)^2 - \frac{\mu_e}{r_0} = \frac{1}{2} v_H^2 \tag{50}$$

Using (41) we can put this in the form

$$(\eta + \chi)^2 = \rho^2 + 2\chi^2 \tag{51}$$

where

$$\chi \equiv \frac{v_0}{v_e} = 0.261786 \tag{52}$$

and

$$\eta \equiv \frac{\Delta v_D}{v_e} \tag{53}$$

$C-2$

(51), together with (41), is the form we will employ in optimization; however, some feeling for the numbers can be had by plotting (50). This is shown in Figure III-3. It may be seen that a few km/sec of v_H cost little more Δv_D than the parabolic escape value, 3.23 km/sec; while for $v_H > 15$ km/sec, we pay for it on essentially a one-for-one-basis.

Implicit in all this is the "patched conic" assumption. That is, the spacecraft is affected by either the earth or the sun, but never both at once. The technique for doing this involves the convenient fiction of the hyperbolic excess velocity, $\overline{v}_H$. Numerous past studies have shown that this yields excellent results for interplanetary missions, and fair accuracy is possible even for circumlunar missions; certainly good enough for this remote solar system feasibility study. For a further discussion of this point, see for instance Reference (9), Chapter 6.

7. A Rough Cut at Spacecraft Design

With the celestial mechanics established, we can turn to the spacecraft. What we will need is a mathematical model for which we can predict the masses of all the important spacecraft components, and particularly the payload, as functions of the optimization parameters Δv_D and Δv_T. Design is a bootstrap process, so we will begin with very crude estimates, and construct a plausible propulsion system.

Our main tool is the rocket equation (see, for instance Reference (36) page 513 et. seq.) which can be stated as:

$$m_0 = m_1 + m_p = m_1 e^{\Delta v/v_c} \tag{54}$$
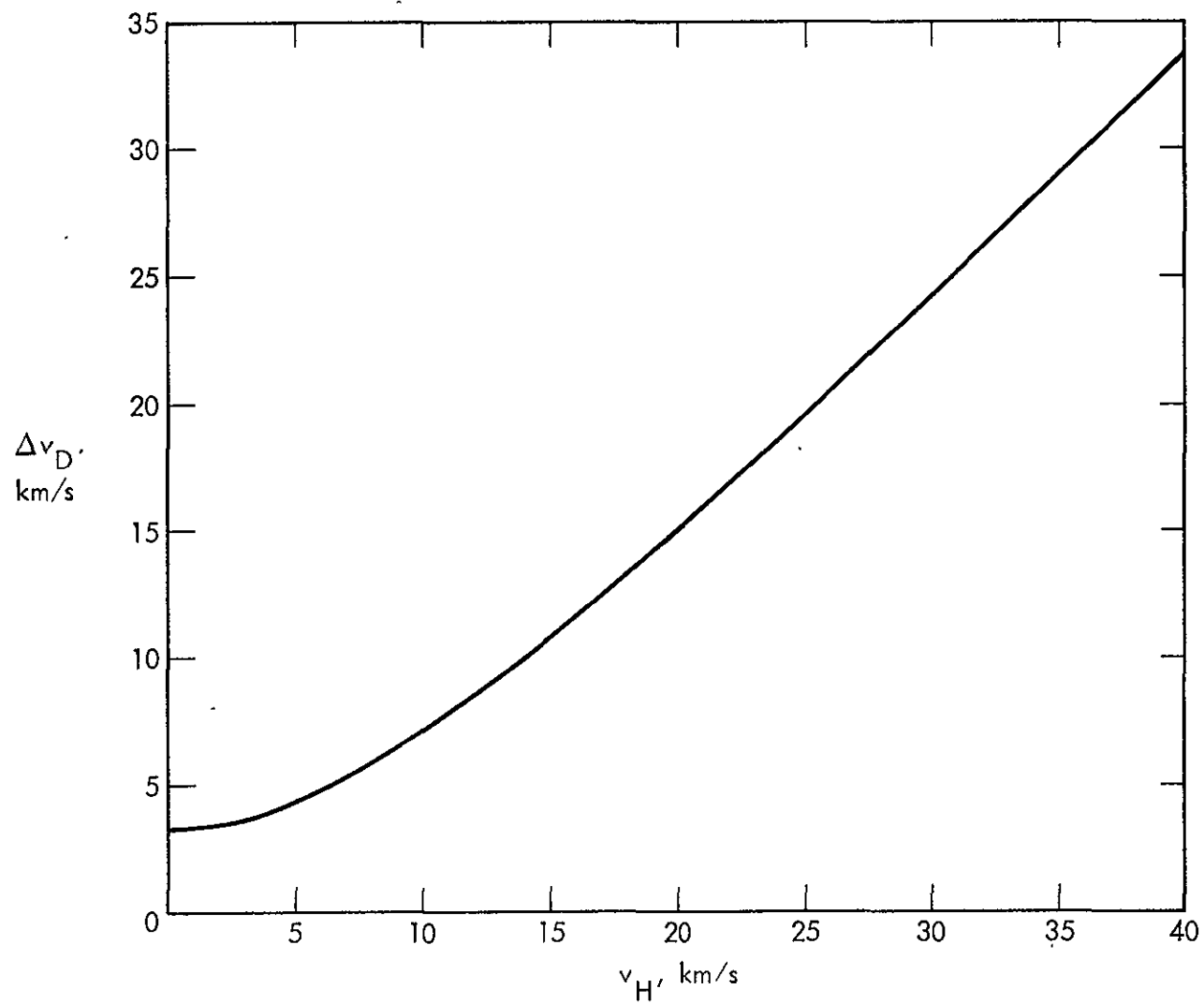
Figure III-3. Δv_D vs Hyperbolic Excess from Low Shuttle Orbit

where m_o is the initial mass of a given stage, m_p is the mass of propellant expelled, and m_l is the mass after losing m_p. The effective exhaust velocity $v_c = I_{sp}$ is a function of the propellants, engine, and nozzle. We can invert (94):

$$m_p = m_o(1 - e^{-\Delta v/v_c}) \qquad (55)$$

Let's look first at departure. In a first crude estimate of Δv_D, we note that we need to reach $r_T \gg r_e$, and have some extra speed when we arrive; but we don't want to overdo it. Parabolic speed $\sqrt{2}\, v_e$ looks about right. Then if we can line up the Shuttle orbit, and for low θ_a, we will need a hyperbolic excess

$$v_H \sim (\sqrt{2} - 1)v_e = 12.3 \text{ km/sec.} \qquad (56)$$

For required impulses of this magnitude it is very important to choose propellants with the highest possible v_c. As long storage is not required for this burn, hydrogen (LH) and oxygen (LOX) are obvious candidates, since they are already used in the Shuttle. To improve on these, we are first led to fluorine (LF), replacing LOX. However, LF must still be regarded as hazardous, and is not likely to be approved as cargo on the man-rated Shuttle, in the foreseeable future. Propellants with higher theoretical performance; e.g. O_3, free radicals, ions; cannot be regarded today as either proven or safe. For these reasons we will choose LH and LOX; but we must show that the low density of LH does not conflict with the Shuttle volume constraints.

This choice of propellants. immediately suggests that we look at the
proven RL-10 engine. The version of this with the most flight experience
is the RL-10A-3-3. Variations of this have been proposed[21] for the Shuttle
Tug and Interim Upper Stage. These are shown in the table below. Because
of volume restrictions in the Shuttle, it was necessary to raise the LOX:LH
mixture ratio, and also fold the nozzle. Eliminating these restrictions
leads to the last column; although we will have to show that we have room
for this. All of these variations have a nominal thrust of

$$T_{RL} = 15,000 \text{ lb} = 66,723 \text{ N}$$

Also, all have provisions for varying the mixture ratio in response to tank
level signals in order to minimize unburnt propellant. However, the level
indicators are not part of the engine.

Table III-1. RL-10 Engine and Derivatives

	RL-10A-3-3	IIB	IV	"NEW"
Mass - kg	131.5	200.5	192.3	176.5
I_{sp} - sec	444	456.6	470	474.5
v_c - km/sec	4.3542	4.4777	4.6091	4.6533
LOX:LH ratio	5.0	6.0	6.0	5.2
Chamber Pressure - psi	400	400	915	915
Expansion Ratio	57	205[1]	401[1]	401
Diameter - m	1.00	1.80	1.80	1.80
Length - m	1.78	1.40 (2.79)[1]	1.40 (2.79)[1]	2.79
Inlet Pressures ⎱ LOX	59.8	19.7	> 10	> 10
(psi) ⎰ LH	38.4	16.4	> 10	> 10
Tank Pressure Source	H_e Tank	Auto[2]	Auto[2]	Auto[2]
Chilldown ⎱ LOX	204.6	THI[3]	THI[3]	64.0
Losses - kg ⎰ LH	40.6	THI[3]	THI[3]	40.6

NOTES: (1) Articulated nozzle fully extended.
(2) O_2 and H_2 gas fed back to tanks through pressure regulators.
(3) Tank head idle feature allows cooldown by burning propellants
    leaking through idle pumps. Very low thrust, but $I_{sp} \sim 375$ sec.
    Also locates ullage.

The source for the first three columns is References (21) and (19). The last column is a variation of the third column, in which the large nozzle is fixed and the THI feature is deleted, in order to save mass; and the mixture ratio is optimized. It was arrived at a discussion[22] with Mr. J. P. B. Cuffe, author of Reference (21). Since a focal mission is unlikely to fly before 1984, engine improvements to raise the chamber pressure, and thus reduce the nozzle mass may be expected. Accordingly, these numbers may be regarded as conservative.

Now let's estimate the overall performance. Assuming the "new" engine, and a Shuttle payload of 29484 kg, and allowing for structure and insulation left behind, and chilldown loss, we will have $m_0 \sim 29{,}000$ kg. Then applying (54) and (55), the burnout mass is $m_1 = 4376$ kg, and the propellant mass is $m_p = 24624$ kg. If we (rather arbitrarily) assume that structure, tankage, and engine amount to 5% of m_p, we end up with 3145 kg at the turn point.

For the second burn we can anticipate a later result that we will need $\Delta v_T \sim 3.2$ km/sec. Considering the long transfer time (about 4 years, as we will see) cryogenic propellants are probably out of the question. Taking $I_{sp} = 300$ sec. as today's practical limit for storable liquids, we get $v_c = 2.9420$ km/sec. This time, the burnout mass is 1060 kg and the propellant is 2085 kg. If we need 8% this time for the dry propulsion system weight, we end up with a payload of 893 kg to explore the focus. Of course, this is only if the axis lies in the ecliptic. As $|\theta_a|$ increases, we can expect our 893 kg to dwindle rapidly.

Now let's examine some side issues. The burn time at departure, t_B, can be obtained from

$$T_{RL} = \dot{m}_p v_c = \frac{m_p v_c}{t_B}$$

(56)

$$t_B = \frac{(24624)(4653.3)}{66723} = 1717.3 \text{ sec.}$$

The RL-10 has been fired continuously[22] for longer than this, so there appears to be no endurance problem. However, this does expose a new difficulty in that during this time the spacecraft would gain considerable altitude, thus violating our assumption of an impulsive Δv_D.

This problem has received considerable attention in the literature. Perhaps best for our purpose is a set of curves of m_0/m_1 vs. I_{sp} at different values of v_H and initial accelerations $n_0 = T_{RL}/m_0 g$, in Ehricke[36]. These were done for tangentially directed thrust, and assumed an initially circular orbit at 500 km altitude. By selecting $I_{sp} = 474.5$ sec, and cross plotting vs. n_0, we are led to Figure III-4. On the right of the figure, Δv_D is obtained from m_0/m_1 by means of (56).

By selecting the curve with $\Delta v_D(\infty) = 8.53$ km/sec, and looking at the "one engine" line, corresponding to a thrust T_{RL}, we can see that to achieve the desired v_H, m_1 is reduced from the impulsive value of 4636 kg to 4000 kg. This large a loss suggests that we try two, or even three, engines to raise n_0. In these cases we find m_1 = 4387 kg and 4531 kg, respectively. From Table III-1 we see immediately that two engines is optimum, even allowing for our lower departure altitude.

There are two reasons for regarding this result as conservative. First, tangential thrust is not optimum, as first shown by Lawden[56,57], although the potential improvement is not large in our case, with two engines.
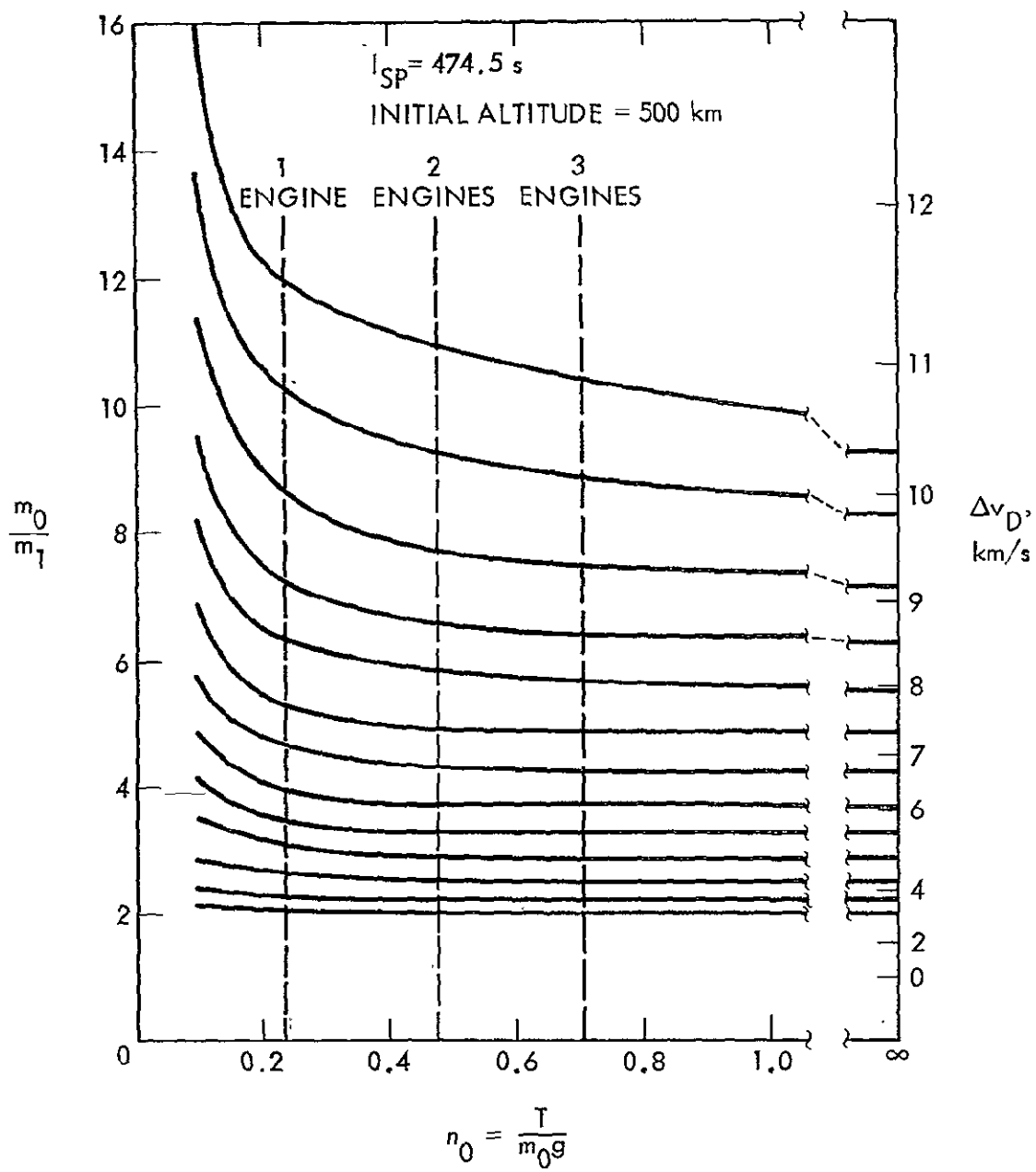
Figure III-4. Gravity Loss due to Finite Thrust

Second, we have the option of applying an initial burn of $\Delta v \sim 1 - 2$ km/sec, leaving us in an eccentric orbit, and completing departure by firing again as we approach perigee. This has the advantage of applying our thrust at a lower average altitude. There is a disadvantage in this of increased propellant boiloff; but this in turn could be minimized by adding the THI feature to the engine (see Table III-1), and operating in this mode whenever the tank pressures got too high. Overall, we can suppose that all these measures might lead to an $m_1 = 4470$ kg. If we put this value back into (54), we find that we can regard the system as impulsive with an effective $v_c = 4.565$ km/sec corresponding to $I_{sp} = 465.5$ sec. For our trial value $\Delta v_D = 8.8$ km/sec, this gives $m_1 = 4219$ kg. We will henceforth adopt these values, as we refine the design.

Now let's look at acceleration. At burnout we get:

$$a = \frac{2T_{RL}}{m_1} = \frac{(2)(66,723)}{4219} = 31.63 \text{ m/sec}^2 = 3.225 \text{ g}$$

This is only slightly higher than the maximum Shuttle acceleration, and it occurs with the departure tanks nearly empty; so it should give us little trouble.

The next question is tank size. From the low tank pressures in Table III-1, we can assume NBP densities[79]:

$$\rho_{LOX} = 1142 \text{ kg/m}^3 ; \quad \rho_{LH} = 70.98 \text{ kg/m}^3$$

With a propellant mass of

$$m_p = m_o - m_1 = 29,000 - 4219 = 24,781 \text{ kg}$$

divided at 5.2:1 we get

$$m_{LOX} = 20{,}784 \text{ kg}; \qquad m_{LH} = 3997 \text{ kg}$$

Allowing, say, 5% for ullage volume, the tank volumes are:

$$V_{LOX} = 19.11 \text{ m}^3; \qquad V_{LH} = 59.13 \text{ m}^3$$

A spherical LOX tank would have a radius r_{LOX} = 1.659 m. However a spherical LH tank would have r_{LH} = 2.417 m, which won't quite fit in the Shuttle bay. The maximum radius[84] is 7.5 ft = 2.286 m; so if we allow the last 15.6 cm for insulation and structure, we get a 2.13 m radius sphere with an inserted cylindrical section 1.308 m long.

The turn point propulsion system will be much smaller, so for rough sizing we can make fairly crude approximations. If we now suppose the dry departure propulsion system at 5.5% of m_p, to allow for the extra engine, we have m_o = 2856 kg at the turn point. Again taking Δv_T = 3.2 km/sec and I_{sp} = 300 sec there we get m_1 = 962 kg, and m_p = 1894 kg. Supposing both propellants have $\rho \sim 1500$ kg/m^3 and mixture ratio 1:1, then after allowing 5% ullage, we get two tanks of volume 0.663 m^3 each. Assuming them spherical, their radii are 0.541 m.

We now have enough to do a rough layout of the spacecraft. Since the LOX tank is by far the heaviest item, structural considerations suggest that it be placed immediately ahead of the engines. Everything else then follows naturally, and we are led to Figure III-5, in which the Shuttle bay outline[84] has been added. Everything appears to fit easily. A few degrees of gimbal freedom for the engines is readily accommodated, and the final

SHUTTLE BAY ENVELOPE
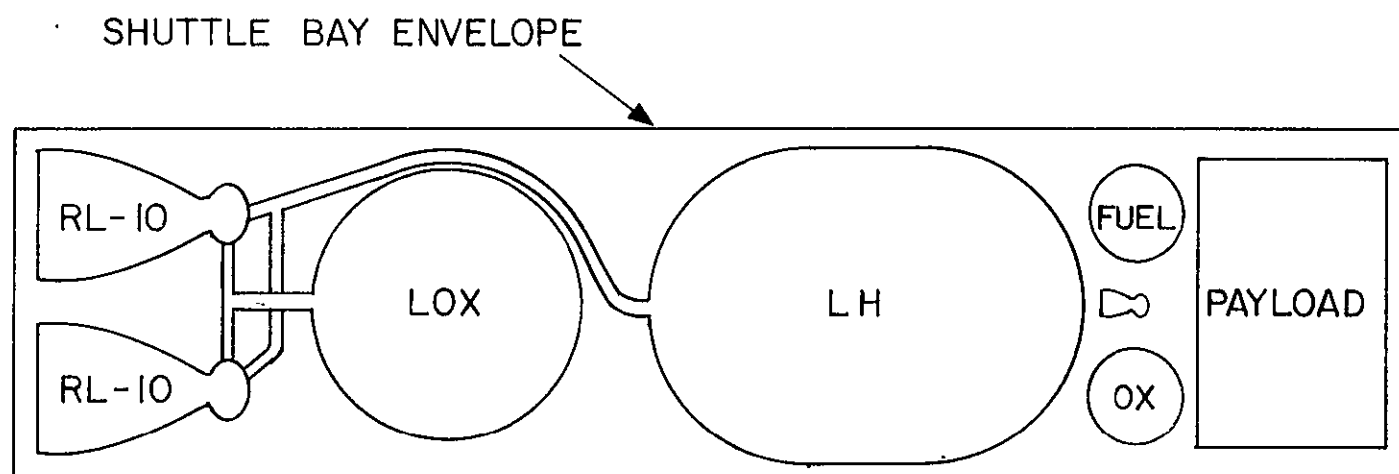
Figure III-5. Focal Probe Configured for Shuttle Launch

payload (about 810 kg in this estimate) should easily fit in the remaining space, even allowing for bulky items such as antennas.

Finally, we need to review our ideas about staging. It has been tacitly assumed that, except for a possible split burn to minimize gravity loss, departure should be achieved by a single stage. Consider, instead, splitting it into two or more Δv impulses; and dropping unneeded engines, tankage, and structure when possible. As we shall show in the next section, the tanks come to only about 70 kg; so it is unlikely that much could be saved by subdividing them. As for structure, Section III-9 shows that it can be built for less than 60 kg; so there is little to be saved there either.

However, from Table III-1, the engines are another matter. Dropping one engine does not look practical, as may be seen from Figure III-5; but jettisoning both, and relying on a single smaller engine, looks feasible. We must compare the potential saving of about 300 kg for part of the burn, with the added masses of the extra engine, plumbing, structure, and dis- connection hardware. A careful design analysis would probably show an advantage to staging, in spite of extra gravity losses, and greater propel- lant residuals and cooldown losses. In the author's opinion, the simplicity of the design of Figure III-5 and the extra Shuttle bay crowding of a staged design far outweigh this performance advantage, and we will not further consider a staged departure. Of course, it is possible that a clever layout and mechanical design could invalidate this argument; so, in this respect, our design is slightly conservative.

8. Propellant Tanks

Let's start with the LOX tank. It is easy to show that a spherical tank of radius r, and filled to a height h, contains a propellant volume

$$V = \frac{\pi}{3} r^3 \theta^2 (3 - \theta) \qquad (57)$$

where

$$\theta \equiv h/r \qquad (58)$$

This is readily checked at the points θ = 0, 1, 2. Supposing the tank to be initially filled to an ullage fraction ε, the initial level θ_0 is given by:

$$\frac{\pi}{3} r^2 \theta_0^2 (3 - \theta_0) = \frac{4\pi r^3}{3(1+\varepsilon)} \qquad (59)$$

Adopting a conservative ε = .05, we get θ_0 = 1.73615.

We can now look at the hydrostatic pressure due to an acceleration a. This is:

$$p_h = \rho h a = \rho r \theta a \qquad (60)$$

During Shuttle launch, the worst a = 3g[84]. The LOX head is then[42]:

$$p_h = (1142)(1.658)(1.73615)(3)(9.80665) = 96712 \text{ N/m}^2 = 14.027 \text{ psi}$$

Things are more complicated during departure. First, V diminishes linearly to zero with time so:

$$\theta^2 (3 - \theta) = \frac{4}{1+\varepsilon} \left(1 - \frac{t}{t_B}\right) \qquad (61)$$

At the same time:

$$a = \frac{2T_{RL}}{m} = \frac{2T_{RL}}{m_1 + m_p(1 - \frac{t}{t_B})} \qquad (62)$$

so, putting these into (60) we get:

$$p_h = \frac{2\rho T_{RL} r\theta}{m_1 + \frac{1}{4}(1 + \epsilon)m_p\theta^2(3 - \theta)} \qquad (63)$$

A plot of this vs. time is shown in Figure III-6. It may be seen that the variation is small till near the end, and that the worst value is

$$p_{hmax} = 15370 \ N/m^2 = 2.23 \ psi$$

The effect of the head pressure depends on how it is passed to the Shuttle structure during launch, and to the RL-10 engines during departure. To reduce tank weight we will assume that some kind of web is used to pass the load to the probe's structure; and that this in turn is anchored to the Shuttle during launch. When this is done, very little of the head pressure appears as stress in the tank walls, and we can design the tank almost solely on the basis of the engine inlet pressure requirement. From Table III-1 we find that we need at least 10 psi, so we can reasonably pick an operating point of 12 psi. Then adding 50% for safety we can design for a yield point of 18 psi. Even during launch this should provide plenty of margin, especially as we will probably subcool the propellants to minimize boiloff.

The tensile stress in a spherical tank of wall thickness w, and containing a pressure p is:

t/t_B
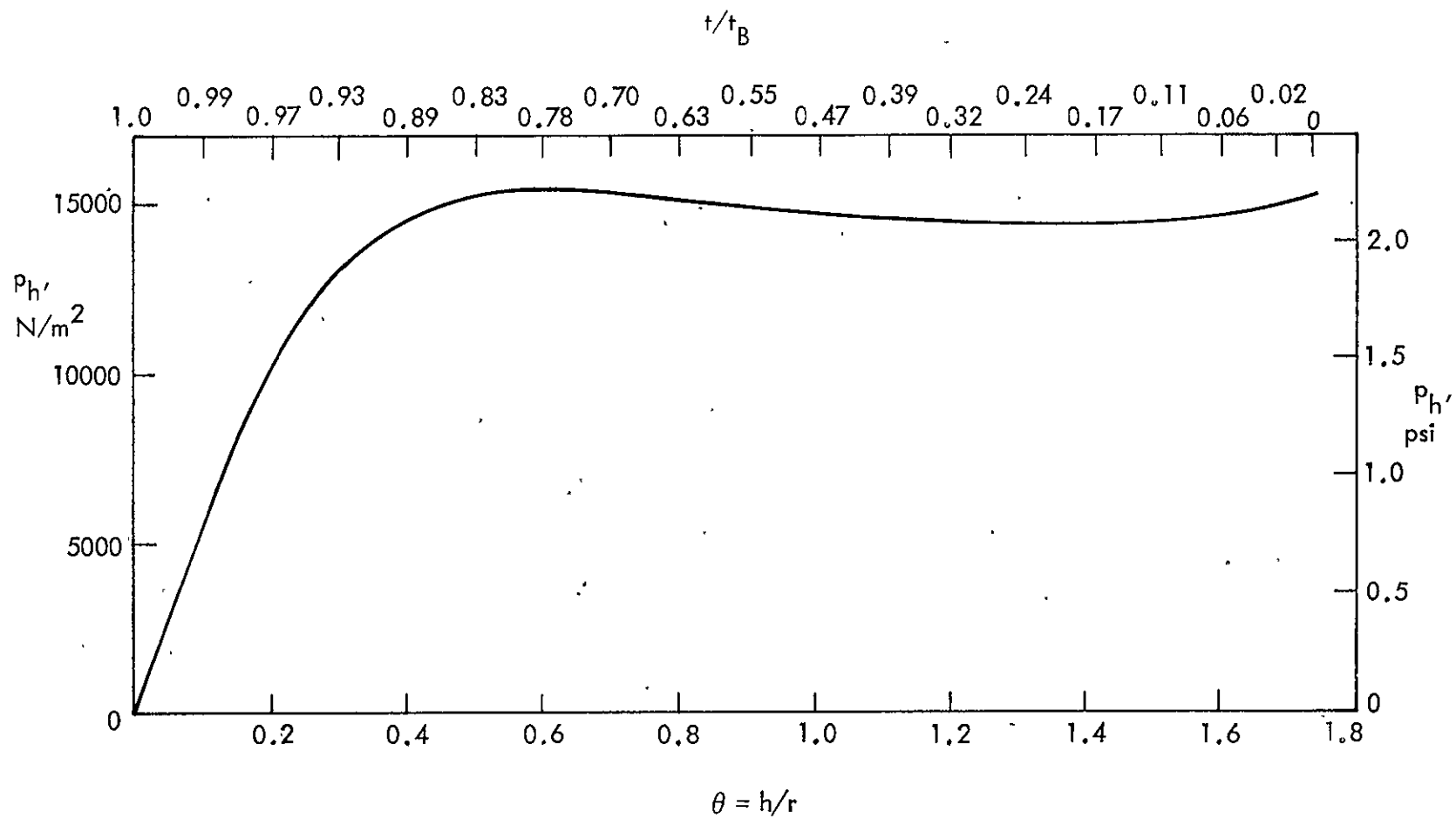
Figure III-6. LOX Hydrostatic Pressure during Departure

$$Y = \frac{rp}{2w} \tag{64}$$

Now, the mass of the tank is

$$m_T = 4\pi r^2 \rho_T w \tag{65}$$

and if Y is the yield stress of the wall material then

$$m_T = \frac{2}{Y}\pi r^3 \rho_T p = \frac{3(1 + \varepsilon)\rho_T p m_{LOX}}{2Y\rho_{LOX}} \tag{66}$$

Thus the tank mass is proportional to the propellant mass, for spherical tanks. All this assumes that the thickness w is above some minimum set by fabrication requirements. Otherwise, m_T is given by (65) with $w = w_{min}$.

Now lets look at some materials. An aluminum alloy recommended for this service is 2219 - T87, for which[55] Y = 67,000 psi at - 320°F. This material has been used for pressure tank service down to LH temperature, and with glue-spot welded joints[86] failed only at ultimate tensile strength. From (64), the required thickness is:

$$w = \frac{rp}{2Y} = \frac{(1.658)(18)}{(2)(67,000)} = 2.23 \times 10^{-4} m = 0.00877 \text{ in.}$$

Actually, fabrication of this tank at thicknesses below about 0.015 in. would be quite difficult. Thus, with a density[61] $\rho_T = 0.103$ lb/in^3 = 2851 kg/m^3, the tank mass from (65) would be:

$$m_T = 4\pi(1.658)^2(2851)(.015)(.0254) = 37.5 \text{ kg}$$

The stronger materials such as titanium and steel are obviously uncompetitive here; but plastic films look very good. Perhaps the best at present is Mylar, with Kapton a close second. Several reports[20,35,46] list $Y \geq 40,000$ psi for some types of Mylar, so in this case w = .0147 in. Moreover, there are no problems in dealing with thin films. Thus, assuming[13] $\rho_T = 1395$ kg/m^3, we find from (66):

$$\frac{m_T}{m_{LOX}} = \frac{(3)(1.05)(1395)(18)}{(2)(40,000)(1142)} = 8.66 \times 10^{-4} \tag{67}$$

which yields $m_T = 18.0$ kg. We will use this ratio in our optimization.

Next we will consider the LH tank. Here, the spherical and cylindrical radius is r = 2.13 m, and the cylindrical length is ℓ = 1.303 m. This time if the tank is filled to a height h, the LH volume is more complicated. If $\theta \leq 1$, (57) is obeyed. When $1 \leq \theta \leq 1 + \phi$ we have

$$V = \frac{\pi}{3} r^3 (3\theta - 1) \tag{68}$$

where

$$\phi \equiv \ell/r (= 0.611737 \text{ with present values}) \tag{69}$$

Finally, when $1 + \phi \leq \theta \leq 2 + \phi$:

$$V = \frac{\pi}{3} r^3 [3\phi + (\theta - \phi)^2 (3 + \phi - \theta)] \tag{70}$$

Using (70), the level θ_0 for an ullage fraction ε obeys:

$$3\phi + (\theta_0 - \phi)^2 (3 + \phi - \theta_0) = \frac{3\phi + 4}{1 + \varepsilon} \tag{71}$$

which for $\varepsilon = .05$ yields $\theta_0 = 2.2896$ and thus $h_0 = 4.8768$ m. The worst hydrostatic head during launch is then (60):

$$p_h = (70.98)(4.8768)(3)(9.80665) = 10184 \text{ N/m}^2 = 1.477 \text{ psi}$$

If we again assume a web support, as in the LOX tank, and some girth support in the cylindrical section from the structure, we can see that this head can be ignored; and of course it will be even smaller during departure. Thus we can again conservatively design the tank to yield at 18 psi.

If for the spherical part of the tank we again look at 2219-T87 aluminum alloy, we find that[55] $Y = 73,000$ psi at -423°F. Thus we would again be forced to minimum thickness, and the mass of the spherical portion would be:

$$m_{TS} = 4\pi(2.13)^2(2851)(.015)(0.2054)$$

However, if we again look at Mylar with[35,46] $Y = 40,000$ psi:

$$w = \frac{(2.13)(18)}{(2)(40,000)} = 4.79 \times 10^{-4} \text{m} = 0.0189 \text{ in.}$$

$$m_{TS} = 4\pi(2.13)^2(1395)(4.74 \times 10^{-4}) = 38.1 \text{ kg}$$

We will adopt this value.

In the cylindrical section, the stress is double that in the end caps, so the design wall thickness is

$$w = \frac{pr}{Y} \tag{72}$$

For Mylar this comes to $w = 9.59 \times 10^{-4}$ in. $= .0377$ in.. For aluminum we get $w = 5.25 \times 10^{-4}$ in. $= .0207$ in., which is now above minimum thickness.

However, the density disparity is so great that $w\rho$ is still lower for Mylar.

Another possibility is titanium. Here[89] Y > 2 x 10^5 psi at $-423°$F for at least two alloys, depending on the stress direction. If we again adopt w = .015 in. as the practical minimum for fabrication and handling, then[61] with ρ_T = 0.161 lb/in^3 = 4456 kg/in^3, we see that titanium does not compete, even with aluminum. Thus, we are again led to choose Mylar. The mass of the cylindrical section with present numbers is then

$$m_{TC} = 2\pi r \ell w \rho_T$$

$$= 2\pi(2.13)(1.303)(9.59 \times 10^{-4})(1345) = 23.3 \text{ kg} \tag{73}$$

More generally, we would like to relate m_{TC} to m_{LH}. For this, we put $\theta = 2 + \phi$ in (70) and get:

$$\frac{\pi}{3} r^3 (3\phi + 4) = (1 + \varepsilon)V_{LH} \tag{74}$$

Putting in (69), and solving for ℓ:

$$\ell = \frac{(1 + \varepsilon)V_{LH}}{\pi r^2} - \frac{4r}{3} \tag{75}$$

Substituting this and (72) into (73) yields:

$$m_{TC} = \frac{2p\rho_T}{Y} \left[\frac{(1 + \varepsilon)m_{LH}}{\rho_{LH}} - \frac{4\pi r^3}{3} \right].$$

$$= \frac{(2)(18)(1395)}{40,000} \left[\frac{(1.05)}{70.98} m_{LH} - \frac{4\pi(2.13)^3}{3} \right]$$

$$= 0.018572 \, m_{LH} - 50.8 \text{ kg.} \tag{76}$$

A final word on the choice of Mylar for the tanks. The author sees no serious disadvantage to its use, beyond the unfamiliarity in this application. However, if problems turn up, and we are forced to aluminum, the nominal cost from the above numbers, for both tanks, is 46 kg, for an ultimate payload loss of about 15 kg.

Finally we will consider the webs supporting the tanks. These consist primarily of a set of straps of length $\frac{1}{2}\pi r$. They must have sufficient combined cross section to sustain the tank weight at the worst launch acceleration, which we will increase to 4g to allow for shock and vibration. Thus the mass of either web is:

$$m_w = (\frac{1}{2}\pi r)(\frac{4gm_p}{Y_w})\rho_w(1.4) = \frac{2.8\pi g\rho_w r m_p}{Y_w} \tag{77}$$

where Y_w is the yield strength of the web material, and we have added 40% to allow for cross straps, tank weight, and plumbing.

An interesting web material is Kevlar, for which[61] $Y_w = 4 \times 10^5$ psi at room temperature, and $\rho_w = 1450$ kg/m^3. This value is for thin filaments, so we should derate Y_w to perhaps half this for weaving into a strap. As Kevlar is likely to be stronger at low temperature, this seems quite conservative. Thus, for the LOX tank web we get:

$$\frac{m_w}{m_{LOX}} = \frac{2.8\pi(9.80665)(1450)(1.658)}{(2 \times 10^5)(6894.7)} = 1.50 \times 10^{-4} \tag{78}$$

in which we have ignored the slight variation of r with m_{LOX}. With the present numbers this is 3.1 kg. For the LH tank, we only need to

change the radius:

$$\frac{m_w}{m_{LH}} = (1.50 \times 10^{-4})(\frac{2.13}{1.658}) = 1.93 \times 10^{-4} \tag{79}$$

which comes to 0.8 kg for the present numbers!

9. Structure

Three main structural systems need to be sized. The primary structure will be a truss to carry the departure load of the two RL-10 engines up to the turn point stage. We will do a fairly detailed synthesis and analysis of this truss. The second stage and payload structure will also be a truss, but as the layout of these components is much less certain, we will make only a crude estimate of the mass in Section 11. The third system is the set of attachments and fittings by which the launch loads are transmitted from the primary and second stage structures to the Shuttle bay structure. We will assume that these connect to the main primary and second stage structural nodes and the RL-10 engines, and are designed to nearly unload those structures during launch.

Referring to Figure III-5 we can see that the primary structure begins just above the RL-10 engines, and ends somewhere around the second stage propulsion system, for a height of about 11 m. The diameter is slightly less than the Shuttle bay, or about 4.5 m. The geometry of the truss is more or less established by the requirement that the ribs lie outside the tanks, and that the only internal tie point lies between the tanks.

These considerations after some experimentation lead to Figure III-7, which shows the main dimensions, and Figure III-8, a photograph of a stick model, showing all the ribs. Most of the departure load is carried by the mainly axial ribs, which come in 3 sets of 10. Letting L be a typical rib length, we define L_1, L_2, and L_3 to be the length of the LOX, intertank, and LH axials respectively. The ribs connecting to the two joints on the axis will be called pyramidals. There are 3 sets of 5 each, and their lengths L_4, L_5, and L_6 refer to the LOX, and lower and upper LH pyramidals respectively. Finally there are the lateral ribs making up the pentagonal cross sections. There are 4 sets of 5 ribs at the axial stations z_1, z_2, z_4, and z_5. These will be called the lower and upper LOX laterals, and lower and upper LH laterals respectively. The length of a lateral rib at z_i will be called ℓ_i. Also, the outer radius of the pentagon at z_i (the dimensions shown in Figure III-7) will be called r_i.

The choice of a pentagonal cross section was by trial and error. An attempt was made to use the simpler square cross section, but it was found that the laterals ℓ_2 were too confining on the LOX tank, and that the axials L_3 were not fully outside the LH radius of 2.13 m. We will reexamine these questions for the present structure.

The lengths of the laterals and pyramidals are easy to compute. For a pentagon:

$$\ell_i = 2r_i s\psi \tag{80}$$

where ψ = 36 deg. The pyramidals obey:

$$L^2 = r^2 + (z_i - z_j)^2 \tag{81}$$
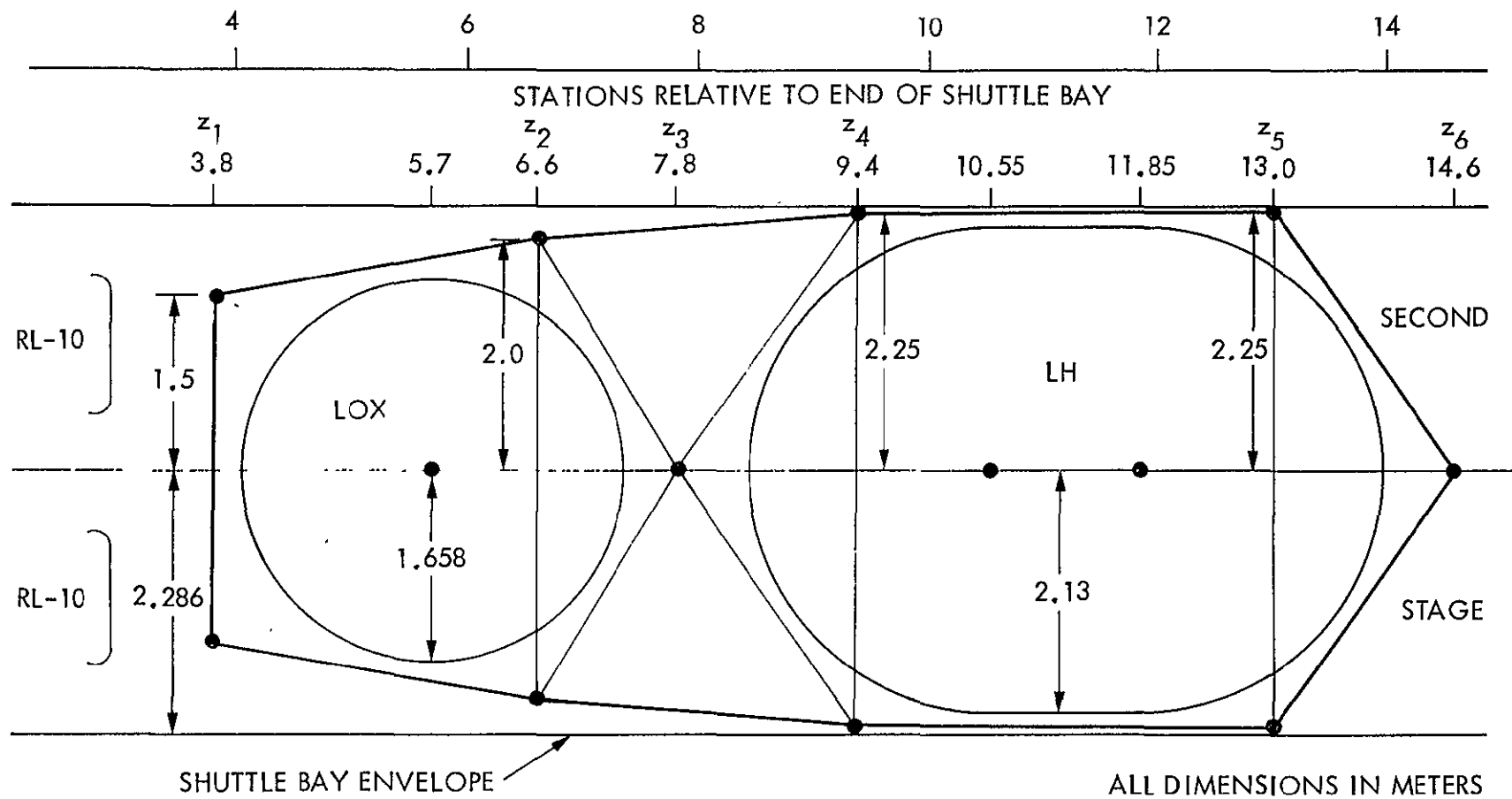
Figure III-7. Primary Structure Geometry
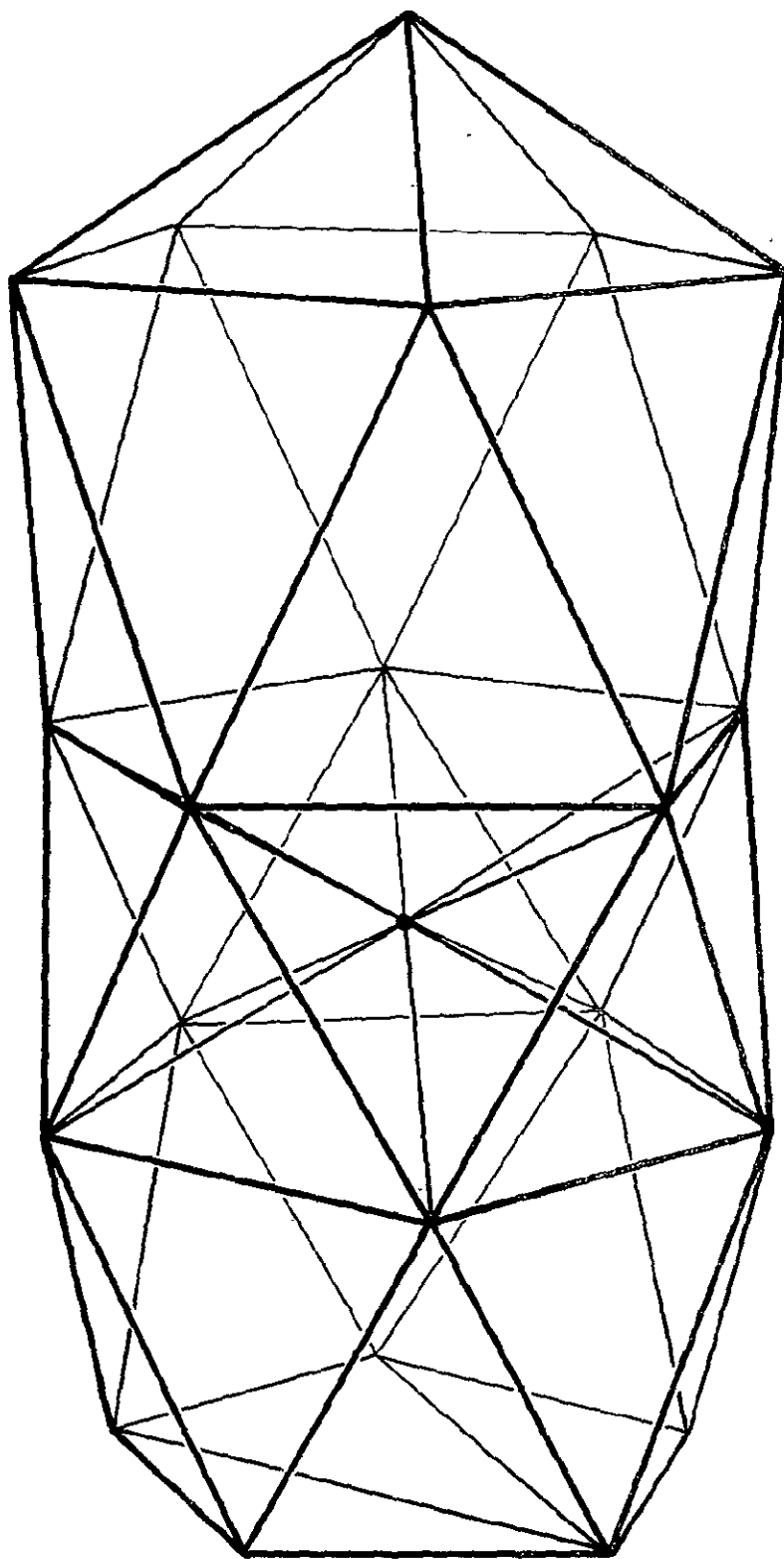
Figure III-8. Stick Model of Primary Structure

Putting in the numbers from Figure 7:

Table III-2.   Truss Dimensions in Meters

INDEX	1	2	4	5
z	3.8	6.6	9.4	13.0
r	1.5	2.0	2.25	2.25
ℓ	1.7634	2.3511	2.6450	2.6450

$L_4 = 2.3324$ m; $L_5 = L_6 = 2.7609$ m

The axials are somewhat harder, since each pentagon is rotated by 36 degrees relative to the one above or below. Suppose the joint at station z_i has cartesian coordinates $(r_i, 0, z_i)$. The joint at the other end of the axial at station z_j will be at $(r_j c\psi, r_j s\psi, z_j)$. Thus the length of that axial is given by

$$L^2 = (r_j c\psi - r_i)^2 + r_j^2 s^2\psi + (z_j - z_i)^2$$
$$= r_i^2 + r_j^2 - 2r_i r_j c\psi + (z_j - z_i)^2 \qquad (82)$$

The results are in Table III-3.

Looking at whether the ribs are too confining, we can see immediately from Figure III-7 that the pyramidals are all right. Looking at the laterals, the mid point of each rib is at a distance D from the point z_c on the axis given by:

$$D^2 = r_i^2 c^2\psi + (z_c - z_i)^2 \qquad (83)$$

For ℓ_2, the distance to the center of the LOX tank is 1.8515 m, leaving a clearance of 19.3 cm. Doing the same for ℓ_4 and the lower LH end cap gives D = 2.1531 m for a clearance of 2.3 cm. Close, but probably adequate. The laterals ℓ_5 are identical to these.

To figure the axials, we can express them parametrically as:

$$x = r_i + \lambda(r_j c\psi - r_i)$$
$$y = \lambda r_j s\psi \qquad\qquad\qquad (84)$$
$$z = z_i + \lambda(z_j - z_i)$$

For L_1, the distance D to the LOX tank center z_c is

$$D^2 = x^2 + y^2 + (z - z_c)^2 \qquad\qquad (85)$$

By differentiation we find that the minimum D requires

$$L_1^2 \lambda = (z_c - z_1)(z_2 - z_1) - r_1(r_2 c\psi - r_1) \qquad\qquad (86)$$

With this, the system is easily solved; and we get

$$\lambda = 0.55684, \qquad D = 1.7309 \text{ m}$$

for a clearance of 7.3 cm.

For L_3 we need the distance to the axis. Since $r_i = r_j = r_4$, and the minimum D obviously occurs at $\lambda = 1/2$ we get:

$$D^2 = x^2 + y^2 = \frac{1}{2} r_4^2 (1 + c\psi) \qquad\qquad (87)$$

From which

$$D = r_4 c(\psi/2) = 2.1399 \text{ m} \quad .$$

for a clearance of just 1.0 cm. Since the spar radius and tank insulation must be allowed for, this is not quite enough; but the necessary slight readjustment of the tank dimensions cannot affect the mass much; so we will accept this and proceed.

The geometry established, we need to compute the load P in each rib. The worst load occurs just at burnout. Assuming, for simplicity, that the entire mass is above the top of the truss at this time, the entire structure is then under an axial compressive load of 2 T_{RL}. This is slightly conservative for the upper ribs, since we have neglected the relief due to the mass of tanks, plumbing, and structure. Again for simplicity, we will assume that the loads are applied symmetrically at the joints at z_1 and z_5.

We will look first at the axials. For L_1 and L_3 we need only divide the load evenly amongst the 10 ribs, and allow for the angle off the axis. This gives

$$P = \frac{T_{RL} L}{5(z_j - z_i)} \tag{88}$$

In the case of L_2, the pyramidals make this part of the structure statically indeterminate. However, these pyramidals are unlikely to be in tension; so it is conservative to compute P_2 from (88). This point would have to be reviewed in a detailed design. The results are given in Table III-3.

The only other important compression load is in the upper pyramidals L_6. If we arbitrarily assume that 1/3 of the second stage load is on the joint

at z_6, and the balance spread uniformly on the joints at z_5, then

$$P_6 = \frac{2T_{RL}L_6}{15(z_6 - z_5)} \qquad (89)$$

Table III-3. Main Compressive Loads

MEMBER	L_1	L_2	L_3	L_6
L (m)	3.0391	3.1018	3.8592	2.7609
P (N)	14484	14783	14306	15351
$LP^{-\frac{1}{2}}$ (m/N$^{\frac{1}{2}}$)	.02525	.02551	.03227	.02228

The calculation of the loads in the laterals is much more slippery; both because the structure is statically indeterminate, and because of the uncertainties in the application of the end loads. Consider a typical outside joint at z_i. From (84), the direction numbers of an axial to z_j are $[r_j c\psi - r_i, r_j s\psi, z_j - z_i]$. At the same joint, the local radial direction is $[1, 0, 0]$. Thus from (82) the angle ϕ between these directions is given by:

$$Lc\phi = r_j c\psi - r_i \qquad (90)$$

Applying this at z_1, we find that $c\phi > 0$; so ℓ_1 is in compression. By resolving all forces at this joint along the radial we learn that

$$P_{\ell_1} = P_1 \frac{c\phi}{s\psi} = \frac{P_1(r_2 c\psi - r_1)}{L_1 s\psi} = 957 \text{ N} \qquad (91)$$

Moving to z_2 and z_4, the problem is indeterminate; but an examination of the model (Figure III-8) suggests that light tension loads are likely in ℓ_2

and ℓ_4; so these members should give us no trouble. As for ℓ_5, the same reasoning as in (90) and (91) this time gives:

$$2P_{\ell_5}s\psi = 2P_3c\phi - P_6 \frac{r_5}{L_6}$$

$$- P_{\ell_5} = \frac{r_4}{s\psi}\left[\frac{P_3}{L_3}(1 - c\psi) + \frac{P_6}{2L_6}\right] = 13352 \text{ N} \qquad (92)$$

This is a tension load, so it's not a major concern. We conclude from all this that the main design problems are the four ribs in Table III-3.

In designing these ribs we will suppose them to be thin walled tubes of radius r and wall thickness w. Three failure modes must be considered - Euler buckling, diamond buckling, and pure compression failure. In the latter case we define failure as exceeding the yield stress. For the design criterion we choose a 100% safety factor; i.e. failure should not occur by any of these modes at less than twice the load of Table III-3.

Starting with compression failure, we have an area limit:

$$\frac{2P}{Y} \leq A = 2\pi rw \qquad (93)$$

Here A is the cross sectional area of the tube, and Y is the yield stress of the tube material. We can write this as:

$$rw \geq \frac{P}{\pi Y} \qquad (CL) \qquad (94)$$

The Euler limit is usually writen:

$$2P \leq \frac{\pi^2 EI}{L^2} = \frac{2\pi^3 Er^3 w}{L^2} \qquad (95)$$

where E is Young's modulus and I is the areal moment of inertia. We will arrange this as:

$$r^3 w \geq \frac{PL^2}{\pi^3 E} \qquad \text{(EL)} \qquad (96)$$

Finally, the diamond buckling limit can be stated as[44]:

$$\frac{2P}{A} \leq \frac{KEw}{r} \qquad (97)$$

where for long tubes $0.1 < K < 0.6$. The higher value is theoretical, and is approachable only for low values of $r/2$, precise shapes, and material nearly free of imperfections. From (93) this takes the form:

$$w^2 \geq \frac{P}{\pi KE} \qquad \text{(DBL)} \qquad (98)$$

It is instructive to sketch these: Two DBL lines are drawn, illustrating the possibilities that the DBL may lie on either side of the point Q_1 where the EL and CL intersect. The inequalities are such that the design point must lie on, or to the right of, all three curves. We note that the EL and CL must lie as drawn, since from (94) the slope of the CL is $-r/w$; while from (96) the slope of the EL is $-r/(3w)$. Thus, the CL is steeper at Q_1.
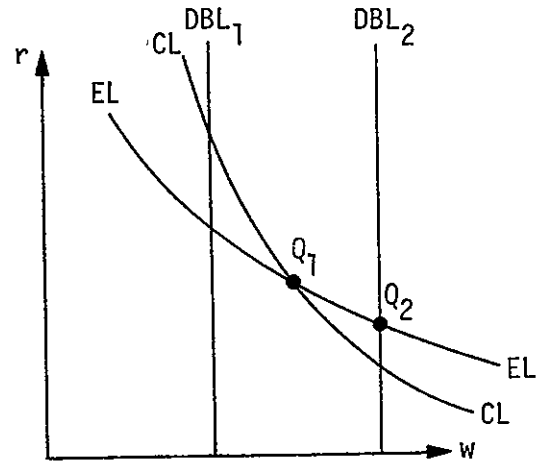
Suppose we have DBL_1. Then since the CL is a contour of constant A, tube mass is minimized by designing to be on the CL somewhere between DBL_1

and Q_1. The tube mass is then:

$$m = \rho LA = \frac{2\rho LP}{Y} \qquad (99)$$

Thin walls are always troublesome; so, to maximize w, we are led to choose Q_1 as the design point. In this case, from (94) and (96) we get:

$$r = \frac{L}{\pi} \sqrt{\frac{Y}{E}} \ ; \qquad w = \frac{P}{L} \sqrt{\frac{E}{Y^3}} \qquad (100)$$

Now consider the converse, DBL_2. In this case, tube mass is clearly minimized at Q_2, the intersection of DBL and EL. Then w is given by (98), and from (96):

$$r^3 = L^2 \sqrt{\frac{KP}{\pi^5 E}} \qquad (101)$$

from which the tube mass is

$$m = 2\rho \left(\frac{L^5 P^2}{\pi K E^2} \right)^{1/3} \qquad (102)$$

To decide between DBL_1 and DBL_2 we can compare w as calculated from (98) and (100). DBL_1 holds provided:

$$\frac{L}{\sqrt{P}} \leq E \sqrt{\frac{\pi K}{Y^3}} \qquad (103)$$

The left side is shown in Table III-3.

Looking at materials, specific alloys of aluminum, titanium, magnesium, steel, and beryllium were considered. In each case E and ρ were fairly

III-47

constant across the range of alloys, except for some with a considerable heavy metal fraction. A selection was made from those alloys suitable for extrusions, and possessing a high value of Y/ρ. Except for beryllium, the information is mainly from Reference (61). The beryllium data is from Reference (12). The main properties E, Y, and ρ are listed in Table III-4, after converting to SI units. The right side of (103) is also computed, in which, for simplicity, we have taken $K = 1/\pi$. This is a sort of bootstrap value, and is subject to later revision.

Table III-4. Material Properties

	Al 7178-T6	Ti 4Al-3Mo-1V (Aged)	Mg ZK60A-T5	Steel 440C	Be XT-40
$E - N/m^2$	7.1704×10^{10}	1.1376×10^{11}	4.4819×10^{10}	1.9994×10^{11}	3.0336×10^{11}
$Y - N/m^2$	5.9156×10^8	1.1514×10^9	$2.0684 \times 10^{8*}$	1.9650×10^9	3.5852×10^8
$\rho - kg/m^3$	2823	4512	1827	7750	1827
$EY^{-3/2} - m/N^{1/2}$	.004984	.002912	.015065	.002295	.044688

*Value is for tubing in compression.

Comparing the last rows of Tables III-3 and 4 we see from (103) that for all ribs, Q_1 is the design point for beryllium, and Q_2 should be used for all other materials. Knowing this, it is a simple matter to apply the above formulas and determine r, w, and m for each rib and material. The results are in Table III-5.

Table III-5. Rib Designs

		Al	Ti	Mg	Steel	Be
m (kg)	L_1	1.2394	1.4563	1.0973	1.7175	0.4486
	L_2	1.2999	1.5273	1.1508	1.8013	0.4673
	L_3	1.8304	2.1507	1.6205	2.5365	0.5627
	L_6	1.0979	1.29	0.972	1.5214	0.432
r (mm)	L_1	109.73	101.61	118.67	92.49	33.26
	L_2	111.62	103.35	120.71	94.08	33.94
	L_3	128.41	118.91	138.88	108.24	42.23
	L_6	103.43	96.24	112.4	87.6	30.21
w (mm)	L_1	0.44944	0.35682	0.5685	0.26915	0.38668
	L_2	0.45406	0.36048	0.57434	0.27191	0.38669
	L_3	0.44667	0.35462	0.565	0.26749	0.30077
	L_6	0.4627	0.36734	0.58527	0.27709	0.45112

It is obvious from Table III-5 that beryllium is best, both from the
lower mass and from the smaller diameter. Moreover, choosing beryllium
spares us from having to justify our choice of K. Further, avoiding Q_2
means that we don't have to worry about interaction of the two buckling
modes, which might give buckling stresses below those we have used.

However, we are not without problems. The values of w are so small
(as little as 0.012 in.) that we may have fabrication and handling problems.
(Considering the high stiffness of beryllium, it is really no worse in this
respect than anything else in the table.) To cure this, and to add more
conservatism, we will increase all w values (and thus all m values) by 50%.

Adding up the masses for all 35 ribs we have just designed, we get a total
of 25.4 kg. This is a saving of 39.9 kg over the next best material,
magnesium. There is also the problem of cost. Beryllium[61] is currently
around $275/kg, not counting fabrication. $10,000 to save 40 kg looks well
spent.

We also need the mass of the remaining ribs. Considering the loads
we have guessed and calculated for these, a good choice would seem to be a
somewhat smaller tube of titanium. Suppose we take r = 25 mm and w = 0.5 mm.
This gives a tensile yield load of 90,430 N, which is very conservative.
Also, from Table III-2, the average length of these ribs is 2.4163 m. At, say
L = 2.5 m, the Euler buckling load is 8818 N, which also looks quite con-
servative. As there are 30 such ribs, their total mass is:

$$m = 30\rho L_{av}A = (30)(4512)(2.4163)(2\pi)(.025)(.0005) = 25.7 \text{ kg}$$

Once again, a shift to beryllium seems warranted. Using the same dimensions,
we now get tensile yield at 28158 N, and Euler buckling at 23515 N, which
still looks good. The mass is now 10.4 kg, a considerable saving.

Some details remain. The 20 outer joints might be 0.5 kg each if
they were primarily magnesium costings, or maybe beryllium, for a total of
10 kg. The two axial joints are probably rings at, say, 1 kg each. There
is a short structure down to the RL-10 engines, which consists of maybe 10
struts of about 1 m each, for a mass of, say, 3 kg. There is structure
connecting to the second stage. This might consist of 5 ribs of about 2 m
each, for about 3 kg more. Finally, there is the disconnection mechanism
for breaking these joints after departure, which will be guessed at 5 kg.
Summing up all these contributions, we find a total primary structure mass
of 58.8 kg.

We are now in a position to look at the attachment system to the Shuttle bay, equipment that stays with the Shuttle. Perhaps the lightest feasible method is just a set of rope connections between the primary structural joints and the bay structure. These could be tightened before launch to put the whole structure in light tension. Since most of these lashings would be hard to get at in orbit, remote line cutters would have to be provided. Allowing say 1 kg per joint for rope, cutter, and cabling, this comes to 20 kg for the primary structure. The second stage and final payload are much more compact, but will probably need more rope. Say 8 kg for this. Next, a separate structure and release will probably be needed to support the RL-10 engines. In view of their large size, 20 kg seems reasonable for this. Finally, we are likely to need a set of specialized attachments for the Shuttle's remote manipulator arm to assist in deployment. Maybe another 15 kg. Putting this all together and allowing 20% for bay attachments and miscellaneous, we find a total attachment mass of 75.6 kg.

10.   First Stage Summary

In this section we will assemble the previous results, examine the remaining vehicle systems, and determine the mass we have left after discarding the first stage, as a function of Δv_D. Our first task is to establish the departure mass m_o. This is the Shuttle payload of 65,000 lb = 29483.5 kg minus those things left in orbit. These consist of the bay attachment system just determined, the chilldown losses, and the insulating shrouds for the LOX and LH tanks.

From Table III-1, the chilldown losses amount to 209.2 kg. If we make use of the two burn maneuver of Section 7, we might need to waste another 100 kg which at an initial Δv of 2 km/sec looks like 154 kg at departure. At Δv_D = 8.8 km/sec, this would cost us 53 kg after departure. This is enough to decide to use the THI feature of the engine. (See Section III-7.) We will thus assume no chilldown loss, but add 30 kg to the combined engine mass to cover the THI equipment.

As for the insulating shrouds, their primary use is to limit heat leak from launch to removal from the Shuttle bay. Once the shrouds are removed, we will count on the initially subcooled propellants, an aluminum reflective coating on the tanks, positioning the Shuttle to act as a sun shield before departure, and the use of THI, as necessary, between departure burns. To estimate the shroud mass, the tank areas are:

$$A_{LOX} = 4\pi r^2 = 4\pi(1.658)^2 = 34.545 \text{ m}^2$$

$$A_{LH} = 2\pi r(2r + \ell) = 2\pi(2.13)[2(2.13) + 1.303] = 74.451 \text{ m}^2$$

If we assume a multiple radiation shield type insulation with, say, 30 layers of 1/4 mil Mylar, each coated with 5 μm of aluminum, we find an areal density of

$$30[(2.5 \times 10^{-4})(.0254)(1395) + (5 \times 10^{-6})(2851)] = 0.6934 \text{ kg/m}^2$$

Allowing 30% for seams, attachments for handling, the slightly larger radius of the shroud, and the lack of a taut fit, we get a total shroud mass of

98.3 kg. We might also note here that a 5 μm aluminum coating on the tanks will add

$$(2851)(34.545 + 74.451)(5 \times 10^{-6}) = 1.5 \text{ kg}$$

to the tank mass. Combining these numbers, our initial departure mass becomes:

$$m_o = 29483.5 - 75.6 - 0 - 98.3 = 29309.6 \text{ kg}$$

On to the first stage. The propellant mass m_p may be calculated as a function of Δv_D from (55). There, m_o is as above, and the effective $v_c = 4.565$ km/sec, as shown in Section III-7. The propellant tank mass m_T can be found by consolidating the various components of Section III-8, including the webs. From (67), (76), (78), and (79):

$$m_{TLOX} = (8.66 \times 10^{-4} + 1.50 \times 10^{-4})m_{LOX} = 1.016 \times 10^{-3}m_{LOX}$$

$$m_{TLH} = 38.1 + (.018572 + 1.93 \times 10^{-4})m_{LH} - 50.8 = .018765m_{LH} - 12.7$$

Using the mixture ratio of 5.2 from Table III-1, and allowing for the aluminum coat:

$$m_T = [(1.016 \times 10^{-3})(5.2) + (.018765)(1)] \frac{m_p}{6.2} - 12.7 + 1.5$$

$$= .0038787 \ m_p - 11.2$$

By making rough estimates of some of the smaller components we can tabulate the first stage systems:

Engines (including THI)	383.0 kg
Primary Structure	58.8
Tanks	$.0038787\ m_p - 11.2$
Engine Actuators, Controls, Cables	20.0
Plumbing and Tank Instrumentation	20.0
Sub Total	$470.6 + .0038787\ m_p$

Adding 5% for miscellany this comes to $494.1 + .0040726\ m_p$.

So far we have made no allowance for propellant residuals. The normal allowance of about 0.5% of m_p would come to around 120 kg. This is so severe as to call for strenuous measures to reduce it. The engines already have mixture controls, and we have allowed for tank level instrumentation. If we augment these with tension gauges in the tank support webs, we should be able to control toward nearly complete exhaustion of both propellants. After allowing for the extra instrumentation, and the slightly reduced performance for deviating from the optimal mixture ratio, we will assume a penalty of 50 kg.

Combining all the above items and using (57), the initial second stage mass comes to:

$$m_2 = m_o - m_p - 494.1 - .0040726 \, m_p - 50$$

$$= 29429.0 \, e^{-\Delta v_D/v_c} - 663.5 \text{ (kg)} \qquad (104)$$

If Δv_D = 8.8 km/sec, this comes to m_2 = 3617.9 kg.

11. The Second Stage

Nearly everything in this stage depends on the choice of propellants. The requirement to store these for several years without refrigeration means that we can rule out anything with a room temperature vapor pressure greater than about 100 psi. Storable oxydizers, by this test, include N_2O_4, H_2O_2, red and white fuming nitric acids, $C\ell F_3$, and $C\ell F_5$. Of these, the last two have promising performance, but are relatively untested. We will regard them as possible improvements, but not sufficiently advanced for conservative design. N_2O_4 has considerable experience, good performance, and appears to be storable indefinitely at modest pressure. H_2O_2 has considerable experience and performance slightly less than N_2O_4; but it is unstable and its decomposition is catalyzed by microscopic impurities. Probably too risky for this mission. Finally, the nitric acids also have lots of experience and good performance. However, while they are normally regarded as storable, holding these corrosives in a light, sealed, untended tank for four years or more could be difficult. All in all, N_2O_4 looks presently hard to beat.

As for fuels, the choice is not so plain. Hydrocarbons, such as RP-1; organics, such as aniline; hydrazine, and its derivatives monomethyl hydrazine (MMH), UDMH, and Aerozine-50 are all storable, and well tested. With the possible exception of hydrazine, which is moderately unstable, all

should work well with N_2O_4 for this mission. A short survey, of modern engines seems to show that MMH is the current choice (Viking Orbiter, Shuttle Orbital Maneuvering System, etc.) and we will go with it too. A summary of propellant properties[13],[96] is useful.

Table III-6. Propellant Characteristics

	MMH	N_2O_4
Density - kg/m^3	874 (25^oC)	1450 (20^oC)
Molecular Weight	46.07	92.02
Melting Point - ^oC	-52.5	-11.2
Normal Boiling Point - ^oC	89.2	21.2
Vapor Pressure at 25^oC - psia	49.6	17.2

We will need the average propellant density $\bar{\rho}$. If the N_2O_4:MMH mass ratio is N, then:

$$m_{MMH} = \frac{m_p}{N+1} ; \qquad m_{N2O4} = \frac{Nm_p}{N+1} \qquad (105)$$

so that

$$\bar{\rho} = \frac{m_p}{V_{MMH} + V_{N2O4}} = m_p\left(\frac{m_{MMH}}{\rho_{MMH}} + \frac{m_{N2O4}}{\rho_{N2O4}}\right)^{-1} = (N + 1)\left(\frac{1}{\rho_{MMH}} + \frac{N}{\rho_{N2O4}}\right)^{-1} \qquad (106)$$

At least two engines are consistent with our requirements - the Rocketdyne RS2101C and the Marquardt R40A. Both are uncooled, pressure fed, and have much experience. Their properties are summarized in Table III-7. The data sources are References 49, 52, 85.

Table III-7. Second Stage Engines

	RS2101C	R40A	R40A (LONG)
Mass - kg	8.2	6.6	18.1*
I_{sp} - sec	292	281	292
v_c - km/sec	2.8635	2.7557	2.8635
N = N_2O_4:MMH ratio	1.512	1.595	1.595
Chamber pressure - psi	115.7		
Expansion Ratio	60	22	60
Diameter - m	0.2723	0.2667	0.4404
Length - m	0.5537	0.4572	1.016
Thrust - N	1334.5	3870.0	4021.4
Inlet Pressures - psi	210	238	238
$\bar{\rho}$ - kg/m^3	1148.6	1156.3	1156.3

*Requires niobium - titanium nozzle extension

Applying (56) to the two versions of the R40A, and using an initial mass of 3617.9 kg, shows that the long engine yields 39.1 kg more payload, even after allowing for the increased engine mass. Comparing this in turn to the RS2101C shows that we get the same performance with a much smaller engine and lower inlet pressure. However, the lower thrust means that the burning time is much longer. Still, unlike departure, the gravity losses are here insignificant; so we will tentatively go with the RS2101C. Again assuming Δv_T = 3.2 km/sec, we find m_p = 2434.5 kg. Thus, from (56), the burn time is:

$$t_B = \frac{m_p v_c}{T} = \frac{(2434.5)(2863.5)}{1334.5} = 5223.8 \text{ sec} = 87.1 \text{ min}$$

This is rather a long burn; however Rocketdyne says[85] that continuous burns as long as 45 min have been run on this engine, with no sign of wear, and no difficulty is foreseen for longer burns. If trouble is found in test, we have the option of shutting down, and restarting when the engine cools. If even this fails, a second engine, or the larger R40A, would cost us 10-12 kg of payload.

Let's look at the propellant tanks. To achieve an inlet pressure of 210 psi, we can add say 10 psi for valve and line losses, and 50% for safety and get a yield point of 330 psi. Assuming that minimum thickness is exceeded here (to be shown), our previous work tells us to look at Mylar[59], titanium[10], and steel[61]. At room temperature:

Table III-8. Tank Materials at Room Temperature

	Mylar	Ti-6A1-4V Aged	Steel 440C
Y - psi	13000	1.6×10^5	2.75×10^5
ρ - kg/m^3	1400	4429	7750

Looking at Y/ρ, we can see that titanium wins narrowly over steel. Moreover, the steel chosen, while possessing a higher Y/ρ than other steels, is not easy to weld, and may have problems with cracking after several pressure cycles. The titanium, on the other hand, is a grade especially recommended for room temperature pressure vessels. Recalling that the propellant tanks run around 0.5 m or larger, we find from (64) the wall thicknesses are roughly

$$w = \frac{rp}{2Y} = \frac{(0.5)(330)}{(2)(1.6 \times 10^5)} = 5.16 \times 10^{-4} \text{ m} = 0.0203 \text{ in.}$$

Minimum thickness is not a problem, so Mylar does not compete here.

If we again allow 5% ullage, then from (66) and (106) the total tank mass is

$$m_T = \frac{3(1 + \varepsilon)\rho_T p}{2Y} (V_{MMH} + V_{N2O4}) = \frac{3(1 + \varepsilon)\rho_T p m_p}{2Y\bar{\rho}}$$

$$= \frac{(3)(1.05)(4429)(330)}{(2)(1.6 \times 10^5)(1148.6)} m_p = .012526 \, m_p \tag{107}$$

For the present numbers this works out to m_T = 30.5 kg.

We also need to consider the pressurization system. The total tank volume is

$$V_T = (1 + \varepsilon) \frac{m_p}{\bar{\rho}} \tag{108}$$

and we must fill this with gas at a pressure p_T = 220 psi. Assuming that we fill this from a gas reservoir of volume V_1 at an initial pressure p_1, which is allowed to blow down to a pressure $p_2 > p_T$ at burnout, then the total gas supply obeys:

$$nRT = p_T V_T + p_2 V_1 \tag{109}$$

Here n is the number of moles of gas, R is the universal gas constant, T is the absolute temperature, and the ideal gas law is assumed adequate up to $p = p_2$.

At the higher p_1, a better approximation is needed, and we will adopt the Van Der Waals equation of state:

$$nRT = \left(p_1 + \frac{an^2}{v_1^2}\right)(V_1 - bn) \tag{110}$$

Introducing the specific volume

$$v \equiv V_1/n \tag{111}$$

we can solve for p_1 and get

$$p_1 = \frac{RT}{v - b} - \frac{a}{v^2} \tag{112}$$

For some candidate gases the constants are[13]

Table III-9. Van Der Waals Constants

	He	H_2	CH_4	Ne	A	CO_2	N_2
a atm - ℓ^2/mol^2	.03412	0.2444	2.253	0.2107	1.345	3.592	1.39
b ℓ/mol	.0237	.02661	.04278	.01709	.03219	.04267	.03913
mol.wt. M	4.003	2.016	16.04	20.18	39.944	44.01	28.02

In Table III-9 we have adopted the units of Reference 13, which requires that we measure V_1 in liters, n in gram-moles, p_1 in atm, where 1 atm = 101,325 N/n^2 14.696 psi, and R = .0826 ℓ-atm/mol-K.

To get the mass of gas tank plus gas we combine (66) and (108) and get

$$m_G = Mn + \frac{3\rho_T p V_1}{2Y} \tag{113}$$

Once again assuming yield at 50% over pressure, we set $p = 1.5\, p_1$. Then using (109) and (111) we can establish a sort of figure of merit F:

III-60

$$F \equiv \frac{m_G}{p_T V_T} = \frac{M + \alpha p_1 v}{RT - p_2 v} \tag{114}$$

in which

$$\alpha \equiv \frac{9 \rho_T}{4Y} \tag{115}$$

Choosing the same titanium alloy, then in the above units this is

$$\alpha = \frac{(9)(4429)(14.696)}{(4)(1.6 \times 10^5)} = 0.91532 \text{ gm/}\ell - \text{atm}$$

We also need to choose p_2. Relative to p_T = 220 psi, a good conservative choice would be about 300 psi = 20.414 atm.

Now if we pick a range of values of v, we can compute a corresponding range of p_1 values from (112), for any given gas. Putting these results into (114) shows that F has a shallow minimum at about p_1 = 180 atm for each of the gases of Table III-9. The results for the best three gases is shown in Table III-10. No others gave F < 2.0.

Table III-10. Gas System Optimization

	He	H_2	CH_4
v - ℓ/mol	0.1555	0.1555	0.115
p_1 - atm	186.32	182.15	172.76
F - gm/ℓ - atm	1.4145	1.2933	1.5257

The clear choice is H_2. This is reinforced by the lower permeability of most materials to H_2, compared to He. Converting to other units we get

p_1 = 2677 psi. Also, combining (114) and (108), and adding 2 kg for the regulator, gas lines, and hardware, we get

$$m_G = 2 + p_T V_T F = 2 + \frac{(1 + \epsilon)p_T F}{\bar{\rho}} m_p \tag{116}$$

$$= 2 + \frac{(1.05)(220)(1.2933)}{(1148.6)(14.696)} m_p = 2 + .017699 \, m_p \quad (kg)$$

With our present numbers, this is 45.1 kg.

This is high enough to look for alternatives. One obvious approach is to pressurize with gas obtained from burning the main propellants, thus eliminating the gas tank. Supposing (optimistically) that we could achieve complete combination to stable end products, a likely result would be

$$5N_2O_4 + 4NH_2NHCH_3 \rightarrow 9N_2 + 4CO_2 + 12H_2O$$
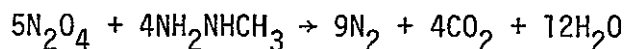
Since the steam would condense at tank temperature, we are left with 13 kg-moles of gas from 644.58 kg of propellants, which gives us an effective M = 49.58 for the gas. Using (108), the mass of extra propellant needed for this is:

$$m_{EP} = Mn = \frac{Mp_T V_T}{RT} = \frac{(1 + \epsilon)Mp_T}{RT \, \bar{\rho}} m_p$$

$$= \frac{(1.05)(49.58)(220)m_p}{(0.0826)(14.696)(300)(1148.6)} = .027383 \, m_p \tag{117}$$

Comparing to (116) shows that this is uncompetitive.

A totally different approach is to pressurize the tanks only lightly, and add a set of pumps to meet the engine inlet pressure requirement. This would greatly reduce both m_T and m_G, at the cost of a set of pumps, and a means to drive them. Let's first consider an electrical drive. From (56), the mass flow rate is

$$\dot{m} = \frac{m_p}{t_B} = \frac{T}{v_c} \tag{118}$$

The theoretical power needed to raise the pressure from nearly zero to p is:

$$P = p\dot{V} = \frac{p\dot{m}}{\rho} = \frac{pT}{\rho v_c}$$

$$= \frac{(220)(1334.5)}{(1.4504 \times 10^{-4})(1148.6)(2863.5)} = 615.4 \text{ W} \tag{119}$$

In view of pump and motor efficiencies, and the various mechanical and pressure losses, it is unlikely that we could get away for less than 1 kW, which certainly makes electrical drive look unattractive.

In place of electrical drive we can consider a gas generator and turbo-pump. Since our main propellants are hypergolic, it is natural to consider them to power the gas generator. To see how much we need, let's look at the main jet. The power developed there is

$$P_E = Tv_c = \frac{m_p v_c^2}{t_B} \tag{120}$$

To drive the turbopump we need the power P from (119). Assuming an efficiency η, the required extra propellant is

$$m_{EP} = \frac{P t_B}{\eta v_c^2} = \frac{P m_p}{\eta v_c T} \qquad (121)$$

Probably no more than 10% efficiency should be assumed for the gas generator and turbopump, but as we will be able to recover some thrust from the exhaust, a value of 20% overall seems more reasonable. Thus:

$$m_{EP} = \frac{(615.4) m_p}{(0.2)(2863.5)(1334.5)} = 8.05 \times 10^{-4} \, m_p \qquad (122)$$

With the present numbers, this comes to m_{EP} = 2.0 kg. Even after we allow say 5 kg for the gas generator, turbopump, and exhaust nozzle, we can see that we are way ahead. A variation on this, suggested by Harold Stratton of Rocketdyne[85], is to power the gas generator by catalytic decomposition of MMH. The combustor may be a little simpler, but the v_c is much less; so m_{EP} would go up by perhaps 2 kg. We will stick with N_2O_4-MMH.

Since m_{EP} is exhausted as we go, rather than remaining as deadweight at burnout, we must determine its effect somewhat differently. Consider the force equation relative to the rocket's initial rest system:

$$m\dot{v} = T = - \dot{m}_p v_c \qquad (123)$$

The mass is here

$$m = m_3 + m_p + m_{EP} \equiv m_3 + (1 + \delta) \, m_p \qquad (124)$$

where we are regarding m, m_p, and m_{EP} as running variables, and $\delta = 8.05 \times 10^{-4}$, from above. Eliminating dt, this integrates to:

$$v + \frac{v_c}{1 + \delta} \ln m = K \qquad (125)$$

and evaluating this before and after burn:

$$\ln(m_2/m_3) = (1 + \delta) \frac{\Delta v}{v_c} \tag{126}$$

For $\delta = 0$, we recognize the rocket equation (54). To make use of the previous theory, we need only replace v_c by

$$v_c' = \frac{v_c}{1 + \delta} = \frac{2.8635}{1.000805} = 2.8612 \text{ km/sec} \tag{127}$$

We are now in a position to recalculate the tank and pressurization system masses. For this we will assume that a tank operating pressure of $p_T = 10$ psi would be sufficient to charge the turbopumps. The mass of the pressurization system is then immediately computed from (116)

$$m_G = 2 + 8.05 \times 10^{-4} m_p$$

The tanks themselves require a new look. Designing them to yield at 15 psi means that both titanium and steel would be well below minimum wall thickness. Even Mylar comes out about .011 in.; but as this does not offer any special problems, it looks like the best choice. From (107), the total tank mass is now:

$$m_T = \frac{(3)(1.05)(1400)(15)}{(2)(13,000)(1148.6)} m_p = .002215 \, m_p \tag{128}$$

which is 5.4 kg with the present numbers.

The second stage structure has to do essentially the same jobs as the first - support the departure loads, and pass the launch loads to the

Shuttle structure. As the payload is poorly defined at this point, we can only roughly estimate its mass. Referring to Figure III-5, and allowing for the diminished departure loads above the second stage tanks, we can guess that the second stage structure will run about 20% of the primary, or 11.8 kg.

We can now tabulate the second stage systems, using a few more guesses for the smaller components:

Engine	8.2 kg
Secondary structure (including tank webs)	11.8
Tanks	$.002215\ m_p$
Pressurization System	$2.0 + 8.05 \times 10^{-4}\ m_p$
Turbopumps, gas generator, nozzle, etc.	5.0
Engine actuators, controls, cables	4.0
Plumbing and tank instrumentation	4.0
Sub Total	$35.0 + .00302\ m_p$

Again allowing 5% for miscellany we are up to $36.8 + .003171\ m_p$. Once again we have to allow for propellant residuals. The usual 0.5% would this time cost us about 12 kg, which without turbopump controls, is likely to be low. However, using the same approach as in the first stage, we can probably get away with 8 kg.

We can now combine all these. Using (55) our payload is

$$L = m_2 - m_p - 36.8 - .003171\ m_p - 8.0$$

$$= m_2\left(1.003171\ e^{-\Delta v_T/v'_c} - .003171\right) - 44.8 \qquad (129)$$

which with the present numbers is 1129.8 kg. Finally, by combining this with (104), we attain our goal of finding the payload as a function of Δv_T and Δv_D only:

$$L = \left(29429\ e^{-\Delta v_D/v_{cD}} - 663.5\right)\left(1.003171\ e^{-\Delta v_T/v_c'} - .003171\right) - 44.8$$

$$= 29522.3\left(e^{-\Delta v_D/v_{cD}} - .022546\right)\left(e^{-\Delta v_T/v_c'} - .003161\right) - 44.8 \qquad (130)$$

12. Optimization

With the construction of the payload function (130) the spacecraft parametric design may be regarded as complete. Our problem now is, given the optic axis coordinates θ_a and ϕ_a, how do we choose trajectory parameters, consistent with the celestial mechanics constraints, which at the same time yield the values of Δv_D and Δv_T that maximize L? This is a straightforward problem in differential calculus, such as is treated by Bryson and Ho (Reference 17, Chapter 1). Our technique will be a variation on their main method.

To start with, we will organize the problem variables into three groups of two each:

$$x \equiv \begin{bmatrix} \sigma \\ \phi \end{bmatrix} ; \qquad y \equiv \begin{bmatrix} n \\ \nu \end{bmatrix} ; \qquad u \equiv \begin{bmatrix} \beta \\ \gamma \end{bmatrix} \qquad (131)$$

where all the Greek variables are defined in Section III-4. This division into "state" variables x and y, and "control" variables u, is rather arbitrary, and is done here merely for convenience.

The celestial mechanics constraints (28) and (31) can be written as:

$$f_1 \equiv \frac{1}{2}[\beta^2 + (1 - \alpha^2)\gamma^2 - \sigma^2] + \alpha - 1 = 0 \qquad (132)$$

$$f_2 \equiv Ae^2 c\phi + (\gamma^2 - 1)(1 - \alpha\gamma^2) - \beta\gamma^2\sigma = 0 \qquad (133)$$

where from (31):

$$e^2 = (\gamma^2 - 1)^2 + \gamma^2\beta^2 \qquad (134)$$

These may be collectively written

$$f(x, u) = [f_1, f_2]^T = 0 \qquad (135)$$

Additional constraints come from the departure relations (41) and (51), and the turn point relation (33):

$$g_1 \equiv \frac{1}{2}(\beta^2 + \gamma^2 - \psi^2 + 1) - \gamma cI + \chi^2 = 0 \qquad (136)$$

$$g_2 \equiv \frac{1}{2}[\alpha^2\gamma^2 + (\sigma - \tau)^2 - \nu^2] = 0 \qquad (137)$$

where cI is from (14) and

$$\psi \equiv \eta + \chi \qquad (138)$$

Collectively these are:

$$g(x, y, u) = [g_1, g_2]^T = 0 \qquad (139)$$

Note that if u is given, (135) may be easily solved for x, after which (139) yields y immediately.

The payload L must also be expressed in these terms. From (130), using (33) and (53) we have:

$$L = L(y) = k_1 \left(e^{-k_5 \eta} - k_2 \right)\left(e^{-k_6 \nu} - k_3 \right) - k_4 \qquad (140)$$

$$k_1 = 29522.3 \text{ kg}, \ k_2 = .022546, \ k_3 = .003161, \ k_4 = 44.8 \text{ kg}$$

$$k_5 = \frac{v_e}{v_{cD}} = \frac{29.78468}{4.565} = 6.5246$$

$$k_6 = \frac{v_e}{v_c^!} = \frac{29.78468}{2.8612} = 10.410$$

To maximize L subject to f = g = 0, we construct a Hamiltonian:

$$H(x, y, u, \lambda, \omega) = L(y) + \lambda f(x, u) + \omega g(x, y, u) \qquad (141)$$

in which λ and ω are Lagrange multipliers with the structure

$$\lambda = [\lambda_1, \lambda_2]; \quad \omega = [\omega_1, \omega_2] \qquad (142)$$

Extremals of the constrained L satisfy these relations:

$$H_x = \lambda f_x + \omega g_x = 0 \qquad (143)$$

$$H_y = L_y + \omega g_y = 0 \qquad (144)$$

$$H_u = \lambda f_u + \omega g_u = 0 \qquad (145)$$

After tabulating all the derivatives of f, g, and L; the system (135), (139), and (143-145) is a set of ten equations in the ten unknowns x, y, u, λ, and ω. We shall devise an iterative technique for solving this system.

First, the derivatives. After some algebra:

$$f_x = - \begin{bmatrix} \sigma & 0 \\ \beta\gamma^2 & Ae^2 s\phi \end{bmatrix}; \qquad f_x^{-1} = \begin{bmatrix} -\dfrac{1}{\sigma} & 0 \\ \dfrac{\beta\gamma^2}{Ae^2\sigma s\phi} & \dfrac{-1}{Ae^2 s\phi} \end{bmatrix} \qquad (146)$$

$$g_x = \begin{bmatrix} 0 & -A(1 - A^2)\gamma Q^{-3} c\phi \\ \sigma - \tau & 0 \end{bmatrix} \qquad (147)$$

$$L_y = - k_1 \left[k_5 e^{-k_5 \eta} (e^{-k_6 \nu} - k_3), \quad k_6 e^{-k_6 \nu}(e^{-k_5 \eta} - k_2) \right] \qquad (148)$$

$$g_y = - \begin{bmatrix} \psi & 0 \\ 0 & \nu \end{bmatrix} \qquad (149)$$

$$f_u = \begin{bmatrix} \beta & (1 - \alpha^2)\gamma \\ \gamma^2(2A\beta c\phi - \sigma) & 2\gamma[A(\beta^2 + 2\gamma^2 - 2)c\phi - \beta\sigma - 2\alpha\gamma^2 + 1 + \alpha] \end{bmatrix} \qquad (150)$$

$$g_u = \begin{bmatrix} \beta & \gamma - cI \\ 0 & \alpha^2\gamma \end{bmatrix} \qquad (151)$$

Now, from (144), we can directly solve for ω:

$$\omega = - L_y g_y^{-1} = - k_1 \left[\frac{k_5}{\psi} e^{-k_5 \eta}\left(e^{-k_6 \nu} - k_3\right), \frac{k_6}{\nu} e^{-k_6 \nu}\left(e^{-k_5 \eta} - k_2\right) \right] \qquad (152)$$

Knowing this, we can obtain λ from (143):

$$\lambda = - \omega g_x f_x^{-1} = \omega \begin{bmatrix} \dfrac{(1 - A^2)\beta\gamma^3 c\phi}{e^2 Q^3 \sigma s\phi} - \dfrac{(1 - A^2)\gamma c\phi}{e^2 Q^3 s\phi} \\ 1 - \dfrac{\tau}{\sigma} & 0 \end{bmatrix} \qquad (153)$$

The combined results must now satisfy (145).

This points the way to the solution technique. A trial value is selected for u, after which x and then y are found from the constraints, as noted above. This leads to ω and $g_x f_x^{-1}$ from (152) and (153). Putting these into (145) gives

$$H_u = \omega(g_x f_x^{-1} f_u + g_u)$$ (154)

In general, this $H_u \neq 0$, and we must decide how to find a better u.

One way to do this is the Newton-Raphson method, which requires the determination of $\frac{d}{du}(H_u)$ to find the change in u. This quantity can be found analytically; but it is evident that the algebra would be extremely tedious. Instead, a numerical technique has been chosen in which H_u is evaluated by the above technique using three different values of u: $[\beta, \gamma]$, $[\beta + \delta, \gamma]$, and $[\beta, \gamma + \delta]$. The results are called H_{u1}, H_{u2}, and H_{u3} respectively. Then

$$\frac{d}{du} H_u = \frac{1}{\delta} \begin{bmatrix} H_{u2} - H_{u1} \\ H_{u3} - H_{u1} \end{bmatrix}$$ (155) .

The improved value of u is then

$$u_{new} = u_{old} - \left[\frac{d}{du} H_u^T\right]^{-1} H_{u1}^T$$ (156)

This procedure is the basis of a set of computer programs, written in APL. The main program is TWOIMP:

```
      ∇ TWOIMP
[1]    C2←2×C1←2□DR×TA
[2]    C3←1-C1×C1.
[3]    GFM← 2 2 ρ0 .
[4]    SW←0
[5]    D←0.01
[6]  Q1:HU3←B IMP G+D
[7]    HU2←(D+D) IMP G
[8]  Q2:HU1←B IMP G
[9]    →SW/Q3
[10]   HU0←((HU2-HU1),[1] HU3-HU1)÷D
[11]   DU←(⍉HU1)⊞⍉HU0
[12]   G←G-W×DU[;1]
[13]   G←G-W×DU[2;1]
[14]   (,HU1),(,DU),G,G
[15]   →SW/Q2
[16]   →Q1
[17] Q3: /PF\SIG,A1×G),SIG,RD×PHI←~2□GF
[18]   (V[ >F[A,NU),ETA,NU
[19]   (V[ >B,G),G,G
[20]   (E←[S*0.5),(AU×GS),RD\~2□CI
[21]   RD\~2□(GSM,~1+A1×GS)÷E
[22]   RHO←FS×0.5
[23]   L←(K1×(K2-*-K5×ETA)×K3-*-K6×NU)-K4
[24]   LQ←(RS+DF×GF←~1+G×C1)*0.5
[25]   P0←~2□GF÷LQ
[26]   CP2←(RHO×G1)-G2(G×0)(CT)÷LQ
[27]   L,(V[ >RHO),RHO,LQ,(RJ×PU),CP2
[28]   →(1-(CP2)/0
[29]   RD×(PHI+P0),PHI←P0+(~P2),P2←~2□CP2
      ∇
```

To calculate H_u for each u, TWOIMP calls a subprogram IMP:

```
      ∇ HU←B IMP G
[1]    SIG←((RS-B×U)+AA+n4×GS←G×G)×0.5
[2]    FS←(GS×BS)+GEM×GSM×GS-1
[3]    CP←QoCP←((U×(S×SIG)+G×n×(A1×GS)-1)×C1)×FS
[4]    [1]←(C1×SP)÷Q(0□CI)×CP
[5]    F←TA×1+PSI←(X2(RS-DS+GS+1-2×G×C1)×0.5
[6]    NU←((A3×GS)+(SIG-TAU)*2)*0.5
[7]    X←K6×(K2-*-K5×ETA)÷HU×K6×NU
[8]    UM← 1 3 ρK1×(K5×(K3-*-K6×NU)-PSI×K5×ETA),(
[9]    GFM[1;1]←~(G×GS÷SIG)□GFM[1;2]←-C5×G×CP÷[S×SP×Q×3
[10]   U□IC,[3]←1-]HU÷SIG
[11]   GUM← 2 2 ρB,(G CI),0,n5/G
[12]   X←~×LG×([J×C1×(R×2)GSM)+GS-B×SIG)×A~×UG
[13]   FUM← 2 2 ρB,n4×G),(GS×(C2×B×CP)-S(U)×X
[14]   HU←U□IC,×GUM×GFM,2[UM
      ∇
```

Finally, to read in some of the parameters, and to calculate constants used in IMP and TWOIMP, there is SETIMP:

```
      ∇ SETIMP
[1]    RD←UK←○÷180
[2]    /X←'ENTER '),'TURN POINT RADIUS IN TM'
[3]    A2←2×A1←(AU÷0.14959797)÷□
[4]    A4←1-A3←A1×A1
[5]    A5←1+A1
[6]    A6←2×A1-1
[7]    Z←'SPEED AFTER TURN IN KM/SEC'
[8]    TAU←□÷VE←29.78468
[9]    Z←'INITIAL ORBIT ALTITUDE IN KM'
[10]   X←Z←(1-□),.19719÷(6371.0.5+□)÷0.5
[11]   R←1.18664
[12]   Q←2.79768
[13]   K1←.79967.5
[14]   K2←0.022546
[15]   K3←0.005161
[16]   K4←44.0
[17]   K5←0.5296
[18]   K6←10.41
      ∇
```

Table III-11. APL Translation

APL	MEANING	APL	MEANING	APL	MEANING	APL	MEANING
AU	r_e	C1	$A = c\theta_a$	GSM	$\gamma^2 - 1$	RD	$180/\pi$
A1	α	C2	$2A$	GUM	g_u	RHO	ρ
A2	2α	C3	$1 - A^2$	HU	H_u	RS	ρ^2
A3	α^2	D	δ	HUU	$\frac{d}{du} H_u$	SIG	σ
A4	$1 - \alpha^2$	DR	$\pi/180$	HU1-3	H_{u1-3}	SP	$s\phi$
A5	$1 + \alpha$	DU	error in u	K1-6	k_{1-6}	SW	switch
A6	$2\alpha - 2$	E	e	L	L	TA	θ_a
B	β	ES	e^2	LQ	q	TAU	τ
BS	β^2	ETA	η	NU	ν	VE	v_e
B1	$s\theta_K/s\varepsilon$	FUM	f_u	OM	ω	W	undershoot
B2	$\cot \varepsilon$	G	γ	PHI	ϕ	X	line break
CI	cI	GF	$\gamma cI - 1$	PSI	ψ	X1	$-\chi$
CP	$c\phi$	GFM	$-g_x f_x^{-1}$	PO	ϕ_0	X2	χ^2
CP2	$c\phi_2$	GS	γ^2	Q	Q		

SETIMP and IMP are easily read, but TWOIMP needs some explanation. To use it we must first run SETIMP, and make an initial guess of u (B, G). Next, the value δ = .01 was chosen as a compromise between too large a step, which would make the derivative approximation (155) poor, and too small, leading to truncation error. This value worked well, so no attempt was made to optimize δ.

The program main loop is [6-16]. After each circuit, the values of H_u, the change in u, and u are displayed. When H_u is sufficiently small, operation is suspended by pushing ATTN. Then the switch SW is set to 1, and execution is resumed at the displayed location. This causes a final calculation of H_u, and the program jumps to [17]. Early use of the program showed that H_u had a tendency to overshoot. Convergence was substantially speeded by introducing a fudge factor W into (156). This was altered as needed between runs, but was generally kept in the range $0.85 \leq W \leq 0.95$.

In the output section, the program first displays many of the variables [17-21]. Then the payload L is calculated, as well as q, ϕ_0, and $c\phi_2$. After displaying these, $c\phi_2$ is examined. If $|c\phi_2| > 1$, the program quits. Otherwise, the limits on ϕ_a, between which the optimal Shuttle orbit can be used, are computed and displayed, along with the center of the excluded region. For I > 0, we find from (11), (40), and the discussion following (49) that the limits are

$$\phi_{a\ell} = \phi + \phi_0 \pm \phi_2 \tag{157}$$

The program assumes this case; but for I < 0 we need only shift $\phi_{a\ell}$ by π.

Most runs were made with r_T = 2.0 Tm, based on reaching the axis well before the focus, according to any plausible solar model of the type considered in Chapter II. Referring to Figures III-1 and 2, we can see that the minimum final specific energy is E ~- 10 km^2/sec^2, which requires v_T ~ 10 km/sec or more. Runs were made with v_T = 10, 12, and 14 km/sec to get a good idea of the payload--time tradeoff. For each of these, a range of values of θ_a up to 19 deg. was tried. The results are shown in Table III-12, and the payload is plotted in Figure III-9. We can conclude from this that payloads well in excess of 1000 kg are possible for $|\theta_a|$ < 5 deg. and v_T < 12 km/sec. Moreover, substantial payload is possible out to $|\theta_a|$ = 15 deg. or so, where the penalty associated with a larger v_T is not so severe.

Everything in Figure III-9 assumes that a perfect line up is possible from an optimal Shuttle orbit. The ecliptic longitude band over which this is possible is also in Table III-12, and is plotted in Figure III-10, for v_T = 10 and 14 km/sec. For handy reference, the celestial equator and a few plausible stellar sources (outward axes) have been added. The loss in Shuttle payload, and thus ultimately in L, for Shuttle orbit inclinations within a few degrees of θ_K = 28.5 deg., is not great; so targets only a few degrees outside the band should be accessible with small loss. Thus for, say, θ_a = 8 deg., ϕ_a = 150 deg., we might get L ~ 500 kg. This is certainly an area for future study.

To see the effect of relaxing our conservative choice of r_T = 2.0 Tm, some runs were made at r_T = 2.5 Tm. Looking again at Figure III-1, we can see that the same range of E_T values is now centered at v_T = 10-12 km/sec, and these two values were examined. The results are shown in Table III-13,
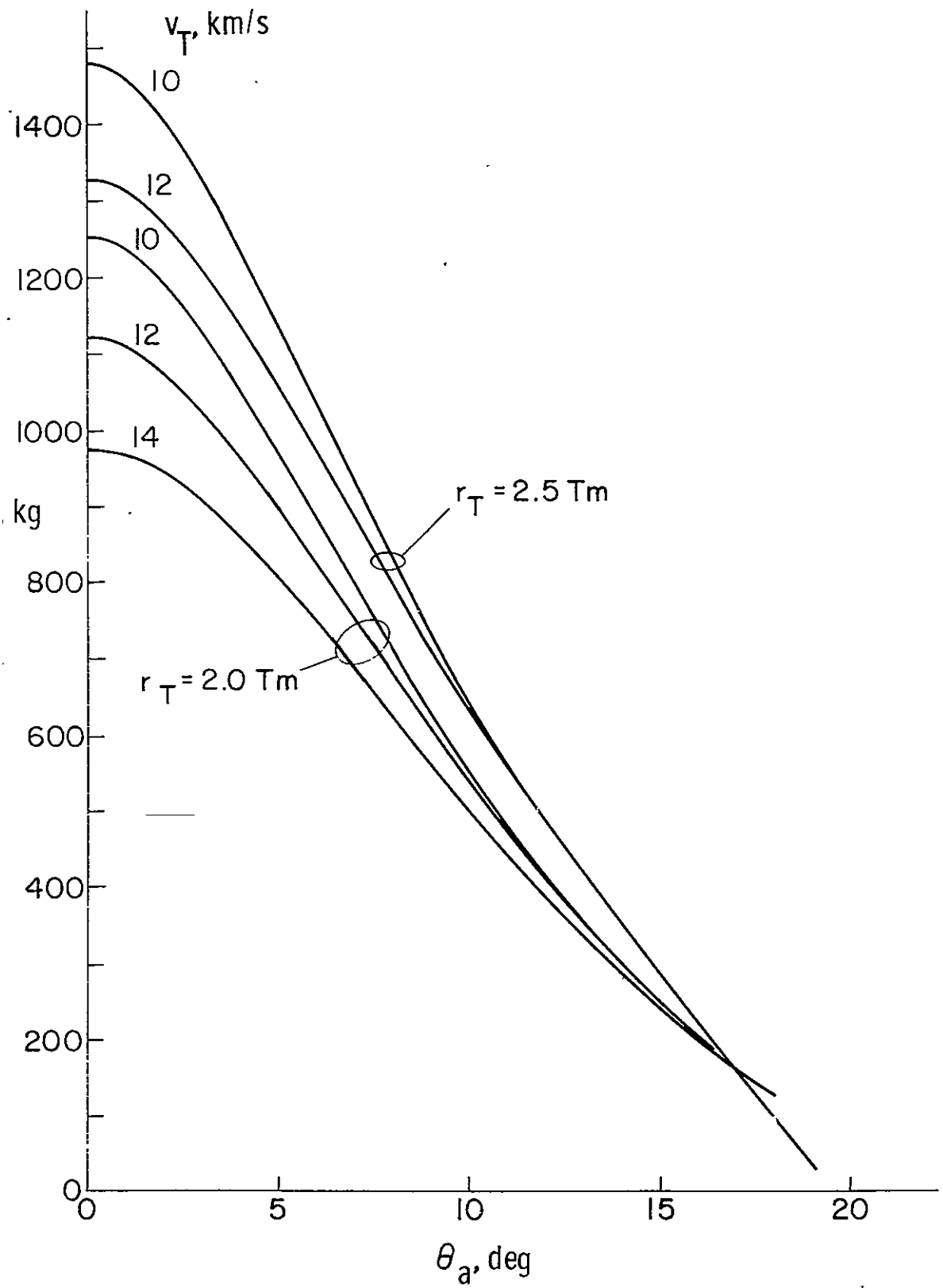
Figure III-9. Final Payload vs Ecliptic Latitude
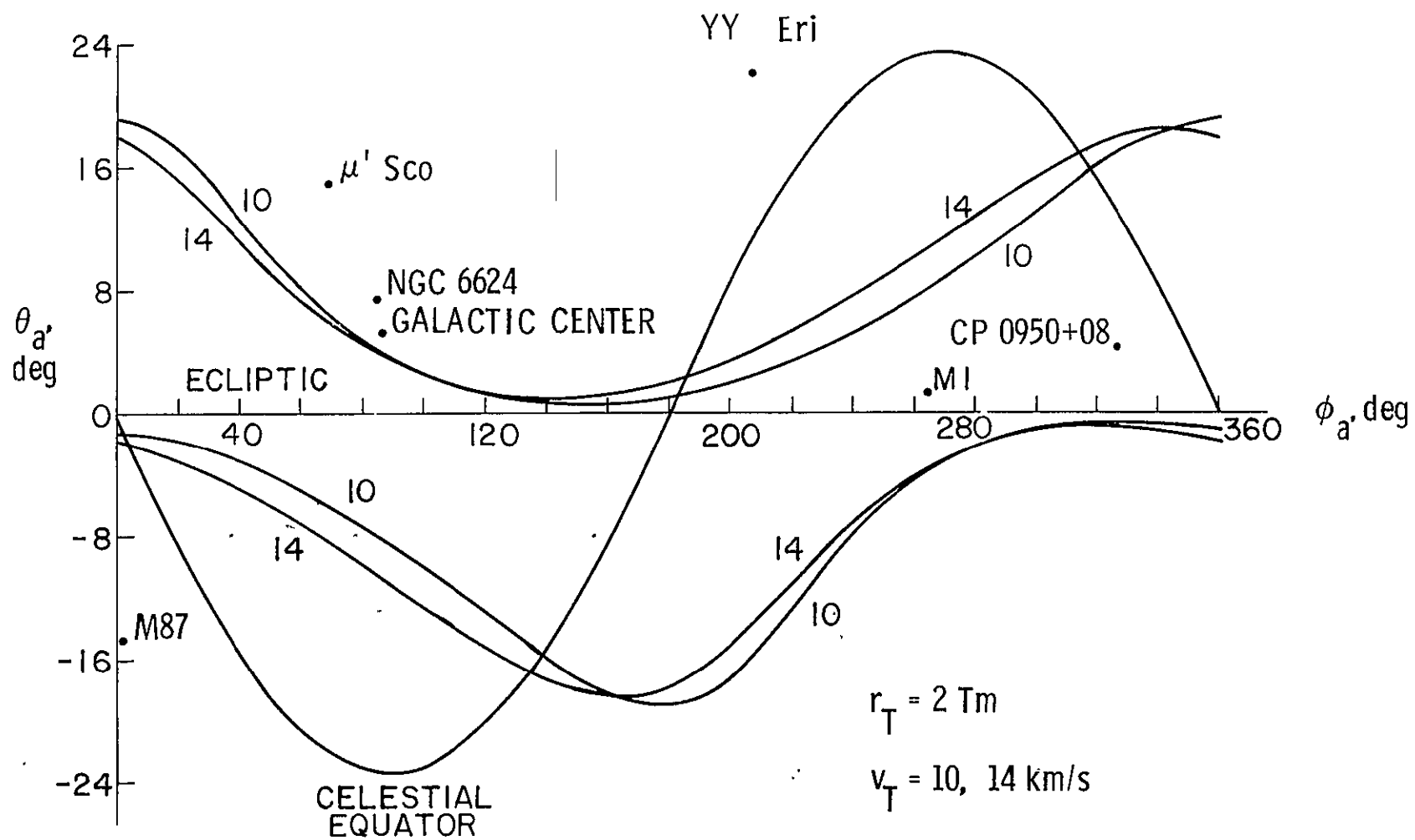
Figure III-10. Optimal Shuttle Orbit Band

Table III-12. Optimal Two Impulse Transfers

θ_a	0	2	5	7	10	12	15	18
Δv_D	8.4708	8.6783	9.4869	10.189	11.368	12.205	13.505	14.834
Δv_T	3.1535	3.1458	3.1276	3.1164	3.0999	3.0888	3.0724	3.0594
v_r	0	0.75623	2.7529	4.0222	5.7499	6.7966	8.2218	9.4804
v_h	41.745	41.753	41.704	41.62	41.44	41.295	41.053	40.807
v_H	11.961	12.241	13.31	14.209	15.675	16.688	18.226	19.763
v_II	9.5589	9.6224	9.7744	9.8586	9.9597	10.017	10.1	10.209
$v_\perp$	3.1225	3.1231	3.1194	3.1132	3.0997	3.0888	3.0707	3.0523
e	0.96439	0.96578	0.9692	0.97116	0.97357	0.97494	0.97695	0.9795
$\ell - Tm$	0.29387	0.29398	0.29329	0.29211	0.28959	0.28756	0.28421	0.28081
I	0	3.9932	8.5325	11.011	14.328	16.379	19.312	22.139
θ_D	0	2.112	7.6735	11.204	16.015	18.933	22.921	26.437
θ_T	152.2	152.04	151.7	151.56	151.45	151.43	151.42	151.35
ϕ	152.2	149.98	144.33	140.88	136.34	133.68	130.12	127
ϕ_0	0	3.6463	13.51	19.969	29.016	34.645	42.544	49.793
ρ	0.40157	0.411	0.44686	0.47707	0.52627	0.56028	0.61192	0.66353
$c\phi_2$	1.1966	0.66954	0.14395	-0.10865	-0.40735	-0.56979	-0.77421	-0.94395
$\phi_{a\ell-}$	--	105.66	76.115	64.608	51.32	43.59	31.934	16.067
$\phi_{a\ell+}$	--	201.6	239.56	257.08	279.4	293.06	313.4	337.52
$L - kg$	1254.8	1190.7	960.96	789.4	552.52	416.87	248.63	118.31

Speeds in km/sec; angles in degrees

$r_T = 2.0$ Tm, $v_T = 10$ km/sec

Table III-12. Continued

θ_a	0	2	5	7	10	12	15	18
Δv_D	8.8247	8.9842	9.6689	10.3	11.397	12.194	13.452	14.758
Δv_T	3.2049	3.1988	3.1804	3.1682	3.1505	3.1388	3.1206	3.1021
v_r	0	0.50848	2.1507	3.3055	4.9413	5.9545	7.3595	8.6287
v_h	42.222	42.231	42.212	42.158	42.023	41.904	41.697	41.468
v_H	12.438	12.65	13.545	14.35	15.711	16.675	18.164	19.676
$v_\parallel$	11.455	11.496	11.619	11.694	11.786	11.835	11.896	11.954
$v_\perp$	3.158	3.1588	3.1574	3.1534	3.1433	3.1344	3.1189	3.1018
e	1.0095	1.0106	1.0138	1.0157	1.0179	1.091	1.0203	1.0214
$\ell - Tm$	0.30062	0.30074	0.30048	0.29971	0.29779	0.29611	0.29319	0.28998
I	0	3.5636	7.988	10.48	13.837	15.918	18.897	21.771
θ_D	0	1.372	5.7937	8.8971	13.294	16.023	19.818	23.259
θ_T	147.32	147.21	146.95	146.83	147.74	146.73	146.77	146.83
ϕ	147.32	145.89	141.43	138.41	134.29	131.82	128.49	125.55
ϕ_0	0	2.355	10.146	15.814	24.154	29.527	37.288	44.679
ρ	0.41758	0.42471	0.45477	0.48179	0.52748	0.55985	0.60984	0.66061
$c\phi_2$	1.1966	0.73543	0.2226	-.0382	-0.35722	0.53612	-0.76732	-0.96376
$\phi_{a\ell-}$	--	105.59	74.438	62.034	47.51	38.923	25.66	5.7049
$\phi_{a\ell+}$	--	190.9	228.71	246.41	269.37	283.76	305.89	334.76
$L - kg$	1120.2	1075.3	895.34	749.32	536.74	410.38	249.63	122.32

Speeds in km/sec, angles in degrees

$$r_T = 2.0 \text{ Tm}, \quad v_T = 12 \text{ km/sec}$$

Table III-12. Concluded

θ_a	0	2	5	7	10	12	15	18
Δv_D	9.2426	9.3703	9.9552	10.522	11.539	12.295	13.505	14.779
Δv_T	3.2668	3.2619	3.2445	3.232	3.2199	3.2023	3.185	3.1681
v_r	0	0.35592	1.6838	2.7115	4.2352	5.2046	6.5734	7.8375
v_h	42.776	42.783	42.781	42.749	42.65	42.554	42.376	42.163
v_H	12.991	13.158	13.912	14.629	15.884	16.796	18.226	19.7
$v_{\parallel}$	13.34	13.368	13.464	13.529	13.609	13.649	13.689	13.699
$v_{\perp}$	3.1996	3.2001	3.2	3.1976	3.1901	3.183	3.1697	3.1538
e	1.0627	1.0634	1.0662	1.068	1.07	1.0708	1.0713	1.0708
$\ell - Tm$	0.30855	0.30866	0.30864	0.30817	0.30674	0.30537	0.30282	0.29978
I	0	3.2319	7.5011	9.9865	13.371	15.478	18.499	21.418
θ_D	0	0.92487	4.3678	7.0274	10.97	13.483	17.044	20.357
θ_T	142.74	142.68	142.48	142.38	142.3	142.31	142.38	142.55
ϕ	142.74	141.8	138.36	135.79	132.12	129.86	126.74	124.07
ϕ_o	0	1.5767	7.5936	12.416	19.886	24.873	32.29	39.621
ρ	0.43616	0.44177	0.4671	0.49115	0.53331	0.5639	0.61193	0.66142
$c\phi_2$	1.1966	0.78841	0.29852	.036155	-0.29521	-0.48649	-0.74053	-0.96301
$\phi_{a\ell-}$	--	105.41	73.322	60.277	44.831	35.624	21.3	-0.675
$\phi_{a\ell+}$	--	181.34	218.58	236.13	259.17	273.84	296.85	328.06
$L - kg$	976.92	944.67	805.53	684.68	499.84	385.89	237.17	116.72

Speeds in km/sec, angles in degrees

$r_T = 2.0$ Tm, $v_T = 14$ km/sec

Table III-13. Optimal Two Impulse Transfers

θ_a	0	2	5	10	15
Δv_D	8.6897	8.8992	9.7174	11.619	13.775·
Δv_T	2.5404	2.5339	2.5186	2.4958	2.4735
v_r	0	0.77231	2.8121	5.8716	8.3957
v_h	42.041	42.046	41.989	41.707	41.301
v_H	12.257	12.537	13.608	15.982	18.541
$v_{\parallel}$	9.6466	9.6994	9.8258	9.9807	10.102
$v_{\perp}$	2.5157	2.516	2.5126	2.4957	2.4714
e	0.99235	0.99351	0.99626	0.99964	1.0022
ℓ - Tm	0.29805	0.29812	0.2973	0.29333	0.28765
I	0	4.0424	8.6311	14.469	19.469
θ_D	0	2.1115	7.6774	16.03	22.956
θ_T	152.57	152.44	152.17	152.01	152.01
ϕ	152.57	150.39	144.8	136.9	130.71
ϕ_0	0	3.635	13.483	28.985	42.523
ρ	0.41151	0.42092	0.45687	0.53657	0.62251
$c\phi_2$	1.1966	0.67195	0.14854	-0.39963	-0.76244
ϕ_a -	--	106.24	76.829	52.327	33.558
ϕ_a +	--	201.8	239.74	279.44	312.92
L - kg	1480.2	1402.9	1126.9	639.79	280.71

Speeds in km/sec, angles in degrees

$r_T = 2.5$ Tm, $v_T = 10$ km/sec

Table III-13. Concluded

θ_a	0	2	5	10	15	19
Δv_D	9.0512	9.2116	9.9032	11.649	13.72	15.479
Δv_T	2.5819	2.5767	2.5612	2.5363	2.5115	2.4914
v_r	0	0.51898	2.196	5.045	7.515 ·	9.2159
v_h	42.523	42.53	42.504	42.299	41.635	42.526
v_H	12.739	12.951	13.846	16.017	18.477	20.499
$v_{\parallel}$	11.562	11.597	11.699	11.838	11.93	12.005
$v_{\perp}$	2.5446	2.5449	2.5434	2.5311	2.5106	2.4914
e	1.0383	1.0392	1.0418	1.0449	1.0464	1.0475
ℓ – Tm	0.30493	0.30502	0.30465	0.30171	0.29683	0.29232
I	0	3.6049	8.0756	13.968	19.045	22.862
θ_D	0	1.3719	5.7965	13.31	19.855	24.388
θ_T	147.74	147.66	147.45	147.3	147.37	147.46
ϕ	147.74	146.34	141.93	134.85	129.08	125.25
ϕ_0	0	2.3473	10.124	24.129	37.274	47.048
ρ	0.4277	0.43481	0.46487	0.53777	0.62035	0.68824
$c\phi_2$	1.1966	0.73766	0.22705	−0.34957	−0.75553	−1.0064
$\phi_{a\ell-}$	--	106.22	75.179	48.522	27.287	--
$\phi_{a\ell+}$	-- ·	191.15	228.93	269.44	305.43	--
L – kg	1323.2	1269	1052.7	624.03	283.66	24.388

Speeds in km/sec, angles in degrees

$r_T = 2.5$ Tm, $v_T = 12$ km/sec

and the L curves have been added to Figure III-9. The optimal Shuttle orbit band looks pretty much the same as in Figure III-10; so it has not been plotted. The large increase in L for $|\theta_a| < 15$ deg tells us that the choice of r_T for the first mission could be very hard. On the other hand, once some focus is found, we will likely be able to carry substantially more L on future missions.

13. Transfer Time

The time along the transfer orbit from departure to the turn point will be called the transfer time. Its calculation from the orbit elements, for any of the orbits just determined, is straightforward; but it is complicated by the need for high precision and two sets of formulas. This is because the orbits are all nearly parabolic; and both ellipses and hyperbolas occur. One way out of this is to merely set e = 1, and calculate the time from the well known formulas for parabolic orbits. This might be accurate enough for our purpose; but an exact treatment for e ~ 1 in terms of power series is not too difficult, and we will develop it here.

The time from perihelion on any Keplerian orbit is:

$$t - t_p = \frac{2A}{\sqrt{\mu\ell}} \tag{158}$$

where A is the area swept out by the orbit, as seen from the sun. Let's consider ellipses first:

$$2A = abM = \ell^2(1 - e^2)^{-3/2}(E - esE) \tag{159}$$

where a and b are the semi-axes of the ellipse, and M and E are the mean and eccentric anomalies. Combining these we get

$$t - t_p = \frac{1}{2} K\left(\frac{\ell}{1 - e^2}\right)^{3/2} (E - esE) \tag{160}$$

where for ℓ in Tm and time in years:

$$K = \frac{2}{\sqrt{\mu}} \left(2\pi\sqrt{\frac{r_e^3}{\mu}}\right)^{-1} = \frac{1}{\pi} r_e^{-3/2} = 5.50126 \tag{161}$$

We will expand (160) by means of a well known connection between E and θ:

$$x \equiv \tan\frac{E}{2} = \sqrt{\frac{1 - e}{1 + e}} \tan\frac{\theta}{2} \tag{162}$$

Thus:

$$E = 2 \tan^{-1}x = 2x(1 - \frac{1}{3}x^2 + \frac{1}{5}x^4 - \frac{1}{7}x^6 + \cdots) \tag{163}$$

and

$$sE = 2s\frac{E}{2} c\frac{E}{2} = 2xc^2\frac{E}{2} = \frac{2x}{1 + x^2} = 2x(1 - x^2 + x^4 - x^6 + \cdots) \tag{164}$$

Using these to eliminate E from (160):

$$t - t_p = K\left(\frac{\ell}{1 - e^2}\right)^{3/2} x[1 - e + (e - \frac{1}{3})x^2 - (e - \frac{1}{5})x^4 + (e - \frac{1}{7})x^6 - \cdots] \tag{165}$$

Finally we introduce what might be called a parabolic anomaly:

$$U \equiv (1 + e)^{-1/2} \tan \frac{\theta}{2} = \frac{x}{\sqrt{1 - e}} \tag{166}$$

$$t - t_p = K(\frac{\ell}{1+e})^{3/2} [U + (e - \frac{1}{3})U^3 + (e - \frac{1}{5})(e - 1)U^5 + (e - \frac{1}{7})(e - 1)^2 U^7 + \cdots] \tag{167}$$

Note that the problems when $e \sim 1$ have disappeared. In fact, if we put $e = 1$, (167) reduces to a well known formula for parabolic orbits.

We need to show that (167) also works for hyperbolas. In place of (159):

$$2A = abM = \ell^2(e^2 - 1)^{-3/2}(eshE - E) \tag{168}$$

while (162)-(164) become:

$$x = \tanh \frac{E}{2} = \sqrt{\frac{e - 1}{e + 1}} \tan \frac{\theta}{2} \tag{169}$$

$$E = 2 \tanh^{-1} x = 2x(1 + \frac{1}{3} x^2 + \frac{1}{5} x^4 + \frac{1}{7} x^6 + \cdots) \tag{170}$$

$$shE = 2 sh \frac{E}{2} ch \frac{E}{2} = 2x ch^2 \frac{E}{2} = \frac{2x}{1 - x^2} = 2x(1 + x^2 + x^4 + x^6 + \cdots) \tag{171}$$

Now, combining (158), (161), (168), (170), and (171)

$$t - t_p = K\left(\frac{\ell}{e^2 - 1}\right)^{3/2} x[e - 1 + (e - \frac{1}{3})x^2 + (e - \frac{1}{5})x^4 + (e - \frac{1}{7})x^6 + \cdots] \tag{172}$$

The parabolic anomaly (166) now becomes

$$U = \frac{x}{\sqrt{e - 1}} \tag{173}$$

with which (172) reduces to (167).

A pair of programs in APL have been written to evaluate (167). These
are:

```
      ∇ TIME
[1]    ZT←TIMEF THT
[2]    ZD←TIMEF TAD
[3]    5.50126÷((SLR÷1+E)*1.5)×ZT-ZD
      ∇

      ∇ Z←TIMEF TA
[1]    US←U×1- (3○DR×TA÷2)÷(1+E)*0.5
[2]    Z←1 (US×E--3
[3]    J←2
[4]    4.X←(E--1(2\J)×((E-1)*J-1)×US*J
[5]    Z←Z+X
[6]    J←J+1
[7]    →(1E-8<|X)/4
[8]    Z←Z×U
      ∇
```

The values of e, ℓ, θ_D, and θ_T are entered, and TIME is called. This
evaluates the portion [····] of (167) for θ_T, and then for θ_D, and computes
the transfer time from the difference. The results, for the orbits
analyzed in the last section, are shown in Figure III-11. By combining this
with Figures III-1 and 2, we can get an excellent idea of the mission
duration.

14. Two Impulse Summary

Our main conclusion from this chapter's work is that starting from a
dedicated Shuttle flight, a focal probe mission near the ecliptic is possible
with a payload in excess of 1000 kg. A new spacecraft would be needed; but
proven engines can be used and no new techniques are called for. Unfortunatel
payload dwindles rapidly as we move away from the ecliptic. About 17 deg.
appears to be the largest excursion from the ecliptic, with any reasonable
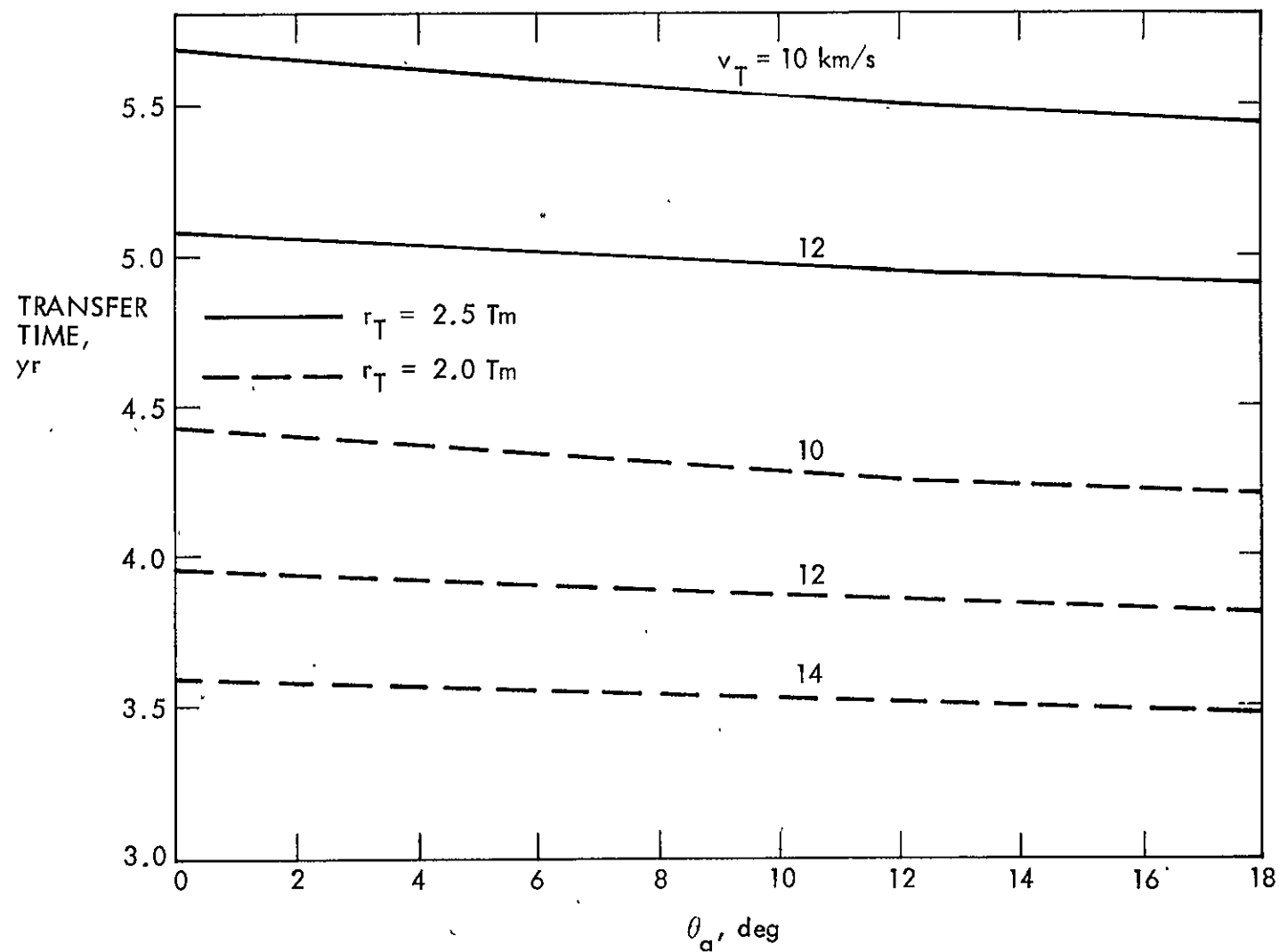
Figure III-11. Transfer Time vs Ecliptic Latitude

payload. Moreover, launch requirements impose severe ecliptic latitude restrictions, outside of which further payload penalties are incurred. In later chapters we shall consider ways to increase payload and extend sky coverage.

For a plausible mission profile, consider the Crab Pulsar (M1). From Figure III-10, we see that it is very favorably located. If we choose the conservative r_T = 2.0 Tm and v_T = 12 km/sec, we get 1100 kg of payload. Transfer time is 4.0 years. If the focus lies in the range of models we have considered (3.36 - 4.43 Tm), then we will encounter it somewhere between 8 and 12 years after launch, and we will have explored out to 5.5 Tm 16 years after launch. If we decide to speed things up by raising v_T to 14 km/sec, then the payload drops to 960 kg, but the four events above now occur at 3.6, 7, 10, and 13 years respectively. This could be a very difficult tradeoff. On the other hand, if we have a good idea of the focal radius, we might choose r_T = 2.5 Tm and v_T = 12 km/sec, which yields a payload of 1300 kg. The four events then occur at 5.7, 8, 11.4, and 15 years respectively. Thus pushing the turn point out considerably increases the payload, but has little effect on mission time.

Finally, a word on the payload composition. We have ignored the impulse requirements of trajectory corrections, for reasons which will become evident later. As for normal spacecraft equipment - electrical power, guidance and control, command and communication, etc.; it is unlikely that any significant fraction of it can be jettisoned before the turn point. Thus, all this must be contained in L, and the actual mass of instruments will be considerably less than L.

CHAPTER IV

ION PROPULSION

1. Introduction

The culmination of the two impulse calculations is that payloads of the order of 1000 kg are feasible for foci near the ecliptic. As the physics instrumentation can only be a fraction of this, and as many interesting potential sources are at high latitudes, improvements should be considered.

One of these, the use of three or more impulses, was mentioned in Section III-1. It does not seem likely that the improvement would be great, at least for reasonable transfer times, so the analysis has not been attempted. Still, this has not been proved; so a valuable resource may have been over-looked.

Another possibility mentioned in Section III-1 is planetary gravity assists, possibly combined with extra impulses. This is an enormously compli-cated subject, partly because of the large number of multiple encounter options, and partly because the optimum departure time and final payload will vary from year to year, due to the motions of the encounter planets. It seems likely that substantial payload increases would ensue, especially at higher latitudes. Someone else's thesis, perhaps.

A different route is to consider radical departures in propulsion. In the chemical area we might, for instance, replace O_2 by F_2 in the first stage, or N_2O_4 by ClF_3 in the second. These were eliminated in the last chapter by the conservative rule of using only tested propellants and engines. We have chosen not to explore these possibilities now, on the grounds that it would be inappropriate in a theoretical thesis. If a better engine-propellant combina-tion turns up, its performance can be readily found by repeating some of the calculations in the last chapter.

In this sense, ion propulsion is a qualitative departure; since its performance limits can be found only by solving a problem in continuous optimal control. In this chapter we will first develop the reasoning leading to the replacement of the second stage by an ion drive, then formally state the combined first stage-electrical power-ion drive optimization as a variational problem. The resulting two point boundary value problem is then solved numerically, for a wide range of two and three dimensional trajectories. The results show very large improvements over two impulse transfers, particularly at higher latitudes.

A different qualitative departure is offered by the solar sail. There has been a substantial reawakening of interest[81] in this old idea, largely because of new engineering developments. These include the fabrication of aluminized Kapton and Kevlar films of the order of 3 μm thick, and methods to deploy and stabilize them.[60,47,27] Pure aluminum films, an order of magnitude thinner, have also been produced[33], but much additional work will be needed to establish their structural and control requirements.

A spacecraft relying purely on solar sail propulsion would have difficulty with trajectory control at great solar distances. Moreover, the payload will almost surely require nuclear electric power. Thus we are led to explore a combination of a sail and an ion drive replacing the second stage. That such an alliance is compatible is not obvious; however we will show that optimal pure ion drive trajectories tend to loiter near the sun for long periods, especially for long transfer times and high source latitudes. This is remarkably similar to optimal sail trajectories to the outer planets and out of ecliptic missions.[77,97]

The author had intended to include such a study in the present work; but time has not permitted it. Considerable engineering work would be needed to define the sail stage mass vs. area, with appropriate allowances for structure, control equipment, and automatic deployment after departure. Then an optimization would be needed, similar to the ion drive work below, but adding sail size and continuous pointing to the other controls. Perhaps, still another thesis.

The ion drive work begins by introducing a nuclear electric power system based on existing technology. Next, a highly simplified analysis is given, indicating that large performance gains are likely from an ion drive, thus motivating a deeper look. As in Chapter III, we then look at axial flight beyond the turn point. This helps to establish a reasonable velocity range at the turn point.

Sections 5-8 set up the main variational problem, leading to a two point boundary value problem of tenth order, with integral side conditions. It is a stationary solution to the optimal trajectory problem, in which the controls are continuous thrust magnitude and direction, the first stage impulse, and the mass devoted to electrical power. The performance index is taken as the turn point mass less the mass of the electrical power system.

All this is relatively easy. Solving the system is not. The last three sections discuss how the problem has been organized for the computer, and the iterative methods employed. The set of computer programs is given, together with the pertinent results. These consist of performance curves as a function of time of flight and of the location of the radiation source. Much of this takes the form of a narrative because many modifications of the programs were needed to get all the results, with partial results emerging at several stages

of this development. The reader interested only in these results could turn directly to the ends of Sections 10 and 11.

2. Electrical Power

During the final phase of the mission, the spacecraft will be so remote that solar power sources would not be very effective. The only plausible alternative seems to be nuclear power. For high power levels fission reactor systems certainly give the best power-mass ratios. Unfortunately, none of these can presently be regarded as state of the art; so we will ignore them. What remains are radioisotope thermoelectric generators (RTG). The state of the art here is surely the systems carried by the Voyager spacecraft, en route to Saturn and the outer planets, and we will build on this technology.

Each spacecraft carries three units, each of which puts out 151 watts (electrical) at launch, and has a mass of 40 kg. Each is self contained in that it requires no additional equipment for thermal control or radiation. The decay rate is such that the power has dropped to 133 watts at Saturn, about 1160 days after launch. All this data was provided by Dr. Jeremy B. Jones of JPL.[53]

For our mission we can reasonably assume that we will need at least as much power as Voyager, so several of these units would be required. In this case, we can conservatively assume that mass is proportional to power output, and define the mass to launch power ratio:

$$\beta = 40/151 = 0.26490 \text{ kg/W}$$

Letting P_D be the power at departure (not significantly different from launch), the mass of this system is:

$$m_R = \beta P_D \qquad (1)$$

Assuming that the available electric power drops exponentially, at least over the life of our mission, then at a time t after departure the power available for propulsion is:

$$P = P_D e^{-t/\tau} - P_B \qquad (2)$$

where P_B is the power reserved for other spacecraft systems. Assuming that the payload and communication systems will need full power only rarely during the trip to the turn point, we will assume P_B = 100 watts as a rough estimate of the average spacecraft power requirement. As this entails only about a 30 kg payload penalty, a better estimate is not justified at this time.

The e-folding time can be found from the above data:

$$\tau = t\left[\ln\left(\frac{P_D}{P+P_B}\right)\right]^{-1} = \frac{1160}{365.25}\left[\ln\left(\frac{151}{133}\right)\right]^{-1} = 25.021 \text{ years}$$

and the half-life corresponding to this is:

$$\tau_H = \tau \ln 2 = 17.343 \text{ years}$$

These figures do not correspond to the isotope used (Pu^{236}, τ_H = 87.6 years)[75], since the drop in thermal efficiency with temperature and the slow poisoning of the thermoelectric elements by radiation play important roles.

3. Ion Propulsion Feasibility

The fact that we are carrying an RTG system with relatively little to do from departure to well beyond the turn point suggests that we put it to work in an electric propulsion system. We can hope that it would substantially reduce or even eliminate Δv_T. Let's look at this.

If we create a plasma at a mass rate q, and exhaust it at a speed v_c, then we get a thrust

$$T = q v_c \qquad (3)$$

If the engine operates at an electrical efficiency η, then the power needed to achieve this thrust is:

$$P = \frac{qv_c^2}{2\eta} = \frac{T^2}{2\eta q} \qquad (4)$$

As ion engine technology is as yet in its infancy, the value we should use for η is quite uncertain. Professor Howard Seifert suggests $\eta = 0.5$ as presently realistic[80], and we will use this throughout. Future developments may allow an upward revision.

Now suppose that of the total power needed for payload and spacecraft operations, a quantity P_E can be diverted most of the time for propulsion. Also, suppose that this is augmented by an additional P_A, solely for propulsion. Then, over a time t, the total extra mass spent on electric propulsion is from (1):

$$\Delta m = qt + \beta P_A \qquad (5)$$

where the decay of the power source has been temporarily ignored.

Supposing that qt is not a large fraction of the average spacecraft mass m, then the velocity increment due to electric propulsion is approximately

$$\Delta v = \frac{Tt}{m} = \frac{t}{m} [2\eta q(P_E + P_A)]^{1/2} \qquad (6)$$

We can now inquire whether there is an optimum tradeoff between q and P_A, which maximizes Δv for a given t and Δm. It is easily shown that (6) possesses an extremal under the constraint (5) provided

$$qt = \beta(P_E + P_A) \qquad (7)$$

That is, the propellant mass should just equal the mass of the power system free for propulsion.

If this condition is met, we find that there is an optimal exhaust speed:

$$v_c^2 = \frac{2n}{q} (P_A + P_E) = \frac{2nt}{\beta} \qquad (8)$$

For example, for 10 years and the above assumptions:

$$v_c = \left[\frac{(2)(0.5)(10)(365.25)(86400)}{0.2649}\right]^{1/2} = 34,515 \text{ m/sec.}$$

Finally, we can combine all these expressions and get

$$\Delta v = \frac{v_c}{2m} (\beta P_E + \Delta m) \qquad (9)$$

For illustration, let m = 3000 kg and P_E = 500 watts, and let an additional 400 kg be devoted to propulsion. Then in 10 years we get

$$\Delta v = \frac{34515}{(2)(3000)} [(0.2649)(500) + 400] = 3063 \text{ m/sec}$$

and the average acceleration is $9.71 \times 10^{-6} \text{ m/sec}^2$.

Well, this argument is not very precise; but it does allow us to conclude that we can replace the second stage with an electric propulsion system and achieve a big increase in payload, although it may cost us some

mission time. Moreover, the RTG need not be grossly larger than that of Voyager. With this as motivation, we will do a much more exacting optimization with the intent of determining payload as a function of axis location and mission time.

4. Assisted Axial Flight

As in Chapter III, we preface our work by considering the one dimensiona problem beyond the turn point. In general this is a much more complicated problem than its predecessor due to the decay of the power source and the los of propellant mass. However, as these are compensating effects, we will neglect them in the interest of obtaining a first estimate of the effect on mission time of adding electrical propulsion.

We start with the equation of motion:

$$\ddot{r} + \frac{\mu}{r^2} = \frac{T}{m} \tag{10}$$

Assuming the right side is a constant, we have an immediate first integral:

$$\dot{r}^2 = \frac{2\mu}{r} + \frac{2Tr}{m} + K \tag{11}$$

where, from the turn point conditions:

$$K = v_T^2 - \frac{2\mu}{r_T} - \frac{2Tr_T}{m} \tag{12}$$

Since time does not appear explicitly, we can express (11) as:

$$t = \int_{r_T}^{r} \left(\frac{2\mu}{x} + \frac{2Tx}{m} + K \right)^{-1/2} dx \tag{13}$$

Of course, for T = 0, this reduces to (III-2).

The integral (13) appears to be resistant to simple analytic treatment; but it is not hard to evaluate numerically. The program IONTIME was written to do this:

```
      ∇ IONTIME
[1]    TMU←2×1.327124E20
[2]    KY←3.168876E‾8×DX←10000000000
[3]    ←1.(Δ'ENTER '),'RT IN TM'
[4]    RT←1000000000000×⎕
[5]    ←2‾2,'VT IN KM/SEC '
[6]    VT ←1000×⎕
[7]    ←3.7,'F/M IN M/SEC2'
[8]    TF←2×⎕
[9]    K←(VT×VT)‾(TF×RT)×(TMU÷RT)
[10]   ←4 ,'RF IN TM'
[11]   RF←1000000000000×⎕
[12]   X←RT×0.05×⍳1+(RF‾RT)÷5
[13]   Y←(K+(TF×X)+TMU÷X)×‾0.5
[14]   M.MM
[15]   J←1
[16]   SUM←0
[17]   ...SUM←SUM,(/TL,1+(1,P,1,1)...,1,1,1)...
[18]   ←(1 J←J+5)/16
[19]   KY×P×SUM
[20]   →0
      ∇
```

This is a simple trapezoidal integrator with a step size $\delta x = 0.1$ Tm, arranged to print out every five steps. The results are shown in Figure IV-1 for $v_T = 10$ km/sec and for various thrust levels in the range $0 \leq T/m \leq 3 \times 10^{-5}$ m/sec^2. The top curve (T = 0) corresponds to about $K = E_T = -15$ km^2/sec^2 in Figures III-1 and 2. Note that the case examined in the last section has $T/m = 9.71 \times 10^{-6}$ m/sec^2, and is very nearly the third curve down. Comparison with Figures III-1 and 2 shows that we get nearly the same performance as for $v_T = 12$ km/sec and T = 0. The lesson is that lower values of v_T are feasible for electrically powered spacecraft.

5. <u>Equations of Motion and Mass</u>

Suppose the electrical propulsion system applies an acceleration $\bar{u}$ to the spacecraft. Then the motion depends purely on $\bar{u}$ and the sun. The equations
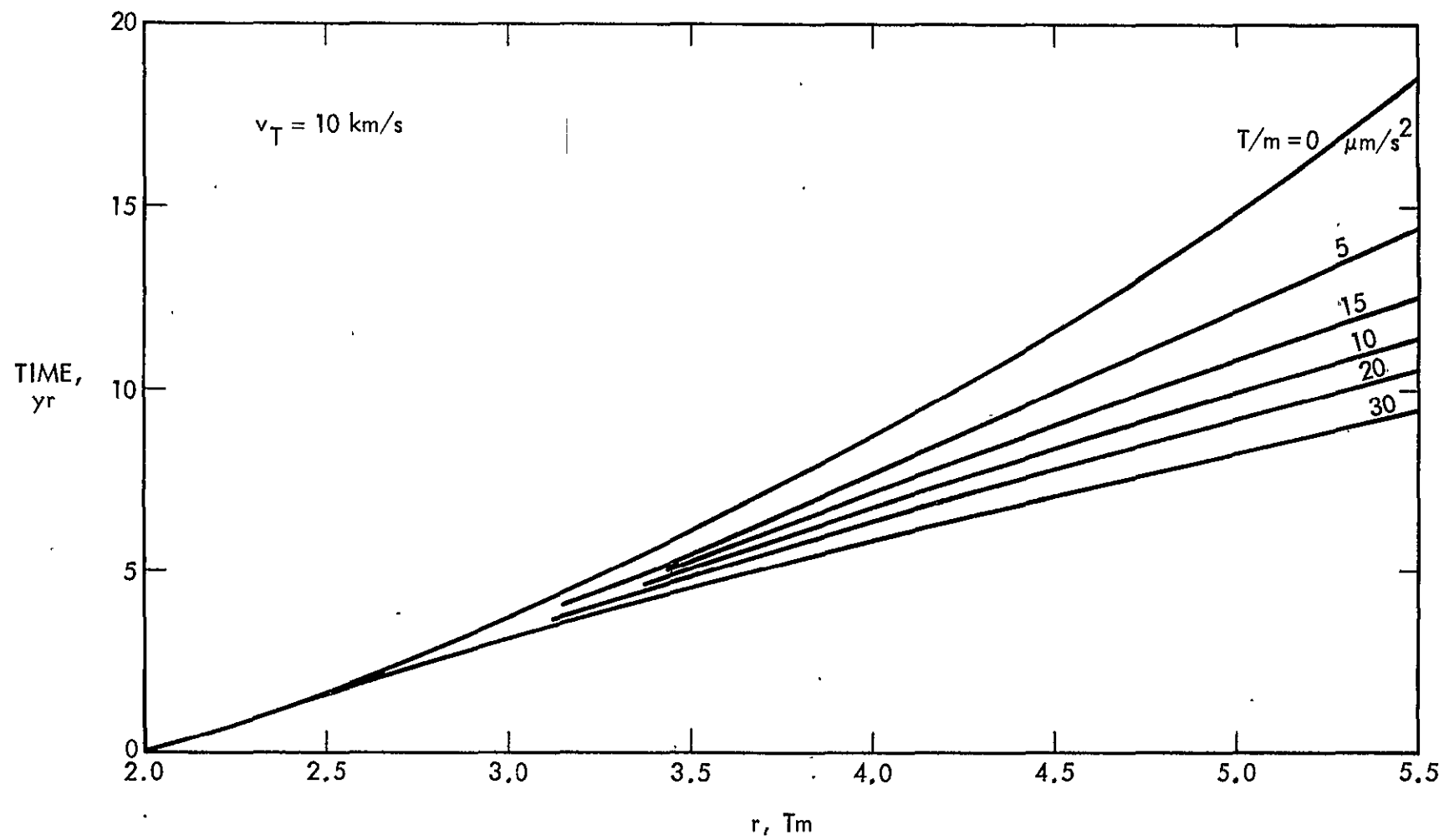
Figure IV-1. Electrically Assisted Axial Flight Time

of motion in first order form are then:

$$\dot{\bar{r}} = \bar{v} \; ; \quad \dot{\bar{v}} = \bar{u} - \frac{\mu}{r^3} \bar{r} \qquad (14)$$

By writing (14) in this form we have been able to suppress the dwindling spacecraft mass m, which would otherwise appear as a state variable. Actually we are only interested in its final value $m_a = m(t_a)$, and its initial value $m_D = m(0)$. These can be related through the propulsion relation (4). Since

$$\bar{u} = \bar{T}/m \qquad (15)$$

we get:

$$\bar{u} \cdot \bar{u} = \frac{T^2}{m^2} = - \frac{2 \eta P \dot{m}}{m^2} \qquad (16)$$

where $\dot{m} = - q$. This integrates to:

$$\frac{1}{m_a} = \frac{1}{m_D} + \int_0^{t_a} \frac{\bar{u} \cdot \bar{u}}{2 \eta P} \, dt \qquad (17)$$

Once the optimal m_D, P_D, and $\bar{u}(t)$ have been determined, we can get P from (2) and evaluate (17) to find m_a.

6. <u>Statement of the Problem</u>

Put loosely, we are out to maximize the useful payload in the focal region, without letting the mission time get too large. The latter is important, as the useful payload probably has no maximum in finite time, although it is clearly bounded above by the Shuttle payload. To avoid this problem, we will specify a terminal time t_a. Also, the need for a turn point has not changed - we still need to reach the axis at $r_T \sim 2\text{Tm}$, with purely axial velocity. However, as we have just seen, the velocity require-

ment is not as great as with two impulse transfer. Putting these thoughts together we can specify a terminal condition:

$$\bar{r}(t_a) = r_T \hat{1}_a; \quad \bar{v}(t_a) = v_T \hat{1}_a \tag{18}$$

where $\hat{1}_a$ is the direction of the axis; and r_T, v_T, and t_a are all given.

The term "useful payload" needs clarification. We will define it as the turn point mass minus the mass of the electrical power system, since the engines themselves have very little mass. We will take this as our performance index. From (1), this index is:

$$J = m_a - \beta P_D \tag{19}$$

The initial conditions also need work. At departure, the earth is at a position $\bar{r}_e$ and has a velocity $\bar{v}_e$, both of which depend only on the departure ecliptic longitude ϕ_D. Thus, the initial conditions are:

$$\bar{r}(0) = r_e \hat{r}_D; \quad \bar{v}(0) = \bar{v}_H + v_e \hat{\phi}_D \tag{20}$$

where $\bar{v}_H$ is the hyperbolic excess velocity from the departure rocket. This is related to the departure impulse through the energy relation (III-50):

$$v_H^2 = \bar{v}_H \cdot \bar{v}_H = (v_0 + \Delta v_D)^2 - 2v_0^2 \tag{21}$$

where v_0 is the (given) Shuttle orbit speed. In turn, the departure mass m_D depends only on Δv_D. Using the first stage design in the last chapter, the relation is (III-104)

$$m_D = k_1 e^{-\Delta v_D / v_c} - k_2 \tag{22}$$

in which the constants have been determined as:

$$k_1 = 29429.0 \text{ kg}, \quad k_2 = 663.5 \text{ kg}$$

$$v_c = 4.565 \text{ km/sec} = 0.1440574 \text{ Tm/year}.$$

Evidently, m_D depends only on $\bar{v}_H$.

We can now state our problem. We wish to start from some condition (20) and reach the state (18) along a path which is consistent with the state equations (14), and the constraint (17). Of all such paths we wish to find one which maximizes the performance index (19), by proper choice of the control $\bar{u}(t)$ and of the initial values $\bar{v}_H$, ϕ_D, and P_D.

7.   Eliminating Longitude

In this form, the standard variational approach leads to a two point boundary value problem of 12th order, with integral path constraints. However, since longitude appears explicitly only in the boundary conditions, it is possible to eliminate it, and thus reduce the problem to 10th order. There _is_ an optimal departure longitude ϕ_D; but we can determine it after the system is solved by integrating $\dot{\phi}$.

Two natural ways of doing this come to mind - cylindrical or spherical coordinates, based on the north ecliptic pole. Of these, cylindricals are perhaps a little simpler. Thus we choose coordinates ρ, ϕ, and z; and resolve all vectors along the orthonormal base vectors $\hat{\rho}$, $\hat{\phi}$, and $\hat{z}$. The first two vary with time, so we need:

$$\dot{\hat{\rho}} = \dot{\phi}\hat{\phi}; \quad \dot{\hat{\phi}} = -\dot{\phi}\hat{\rho}; \quad \dot{\hat{z}} = 0 \tag{23}$$

Resolving $\bar{r}$ and $\bar{v}$ in these coordinates we have

$$\bar{r} = \rho\hat{\rho} + z\hat{z}; \quad \bar{v} = v_\rho\hat{\rho} + v_\phi\hat{\phi} + v_z\hat{z} \tag{24}$$

and combining these with the equations of motion (14):

$$\dot{\rho} = v_\rho, \quad \rho\dot{\phi} = v_\phi, \quad \dot{z} = v_z$$

$$\dot{v}_\rho = u_\rho - \frac{\mu\rho}{r^3} + \dot{\phi}v_\phi, \quad \dot{v}_\phi = u_\phi - \dot{\phi}v_\rho, \quad \dot{v}_z = u_z - \frac{\mu z}{r^3}$$

$$\tag{25}$$

Now we choose state variables:

$$x^T = [\rho, \ v_\rho, \ v_\phi, \ z, \ v_z] \tag{26}$$

In terms of these we can eliminate $\dot{\phi}$ and write (25) in the form

$$\dot{x} = f(x, \overline{u}) \tag{27}$$

where

$$f^T = \left[v_\rho, \ u_\rho - \frac{\mu\rho}{r^3} + \frac{v_\phi^2}{\rho}, \ u_\phi - \frac{v_\rho v_\phi}{\rho}, \ v_z, \ u_z - \frac{\mu z}{r^3} \right] \tag{28}$$

We can now see that we have achieved our aim. Since

$$r^2 = \rho^2 + z^2 \tag{29}$$

it is evident that the system (27) is complete, and solvable once $\overline{u}$ is given. The boundary conditions are easy to write. At the turn point:

$$x^T(t_a) = [r_T c\theta_a, \ v_T c\theta_a, \ 0, \ r_T s\theta_a, \ v_T s\theta_a] \tag{30}$$

and at departure:

$$x^T(0) = [r_e, \ v_{H\rho}, \ v_e + v_{H\phi}, \ 0, \ v_{Hz}] \tag{31}$$

8. Stationary Paths

The standard procedure in this kind of optimization is to attach the state equations to the performance index by Lagrange multipliers. Thus, from (27), we define

$$J^* = J + \int_0^{t_a} \lambda(t)(f - \dot{x})dt \tag{32}$$

where $\lambda(t)$ is a five element row vector that we are free to choose. After integrating by parts this becomes:

$$J^* = m_a - \beta P_D - [\lambda x]_0^{t_a} + \int_0^{t_a} (\lambda f + \dot{\lambda}x)dt \tag{33}$$

On physical grounds, we would expect that some choice of $\bar{u}(t)$, $\bar{v}_H$, and P_D would maximize J^*, and thus J. Since no bounds have been given for any of these controls, we would expect that at the optimum, J^* will be stationary with respect to arbitrary infinitesimal variations of all the controls. Finding such a combination does not guarantee a maximum; but if it gives reasonable values of the parameters we can argue that at least the computed value of J is practical; and this is sufficient to establish a minimum feasibility level. Thus we need to look for paths satisfying $\delta J^* = 0$ in spite of independent arbitrary control variations $\delta\bar{u}$, $\delta\bar{v}_H$, and δP_D; and the resulting state variations δx.

First we will tabulate the variations of some intermediate parameters. From (21) and (22) we get:

$$\delta m_D = (\partial_{\bar{v}_H} m_D) \cdot \delta\bar{v}_H = - \frac{m_D + k_2}{v_c(v_o + \Delta v_D)} \bar{v}_H \cdot \delta\bar{v}_H \tag{34}$$

Also, from (17) and (2) we can compute:

$$\delta m_a = \left(\frac{m_a}{m_D}\right)^2 \delta m_D + \frac{m_a^2}{\eta} \int_0^{t_a} \bar{u} \cdot \left(\frac{P + P_B}{2 P_D P^2} \bar{u} \delta P_D - \frac{1}{P} \delta \bar{u}\right) dt \qquad (35)$$

Since the terminal state (30) is fixed,

$$\delta x(t_a) = 0 \qquad (36)$$

On the other hand, at departure:

$$\delta x^T(0) = [0, \ \delta v_{H\rho}, \ \delta v_{H\phi}, \ 0, \ \delta v_{Hz}] \qquad (37)$$

We are now ready to vary J^* in (33). With the aid of (34-7) we get:

$$\delta J^* = \bar{\lambda}_{235}(0) \cdot \delta \bar{v}_H - \frac{m_a^2(m_D + k_2)}{v_c m_D^2(v_0 + \Delta v_D)} \bar{v}_H \cdot \delta \bar{v}_H - \beta \delta P_D + \int_0^{t_a} \Big[(\lambda f_x + \dot{\lambda})\delta x$$

$$+ \lambda f_u \delta \bar{u} - \frac{m_a^2}{\eta P} \bar{u} \cdot \delta \bar{u} + \frac{m_a^2(P + P_B)}{2\eta P_D P^2} \bar{u} \cdot \bar{u} \delta P_D \Big] dt \qquad (38)$$

where

$$\bar{\lambda}_{235} \equiv [\lambda_2, \ \lambda_3, \ \lambda_5] \qquad (39)$$

At this point, the standard procedure is to pick λ in such a way as to eliminate the awkward δx terms. Also, we will find it convenient to define[*]

$$\lambda = \frac{m_a^2}{\eta} \omega \qquad (40)$$

---

[*]Suggested by J. V. Breakwell.

Thus, the modified multipliers obey

$$\dot{\omega} + \omega f_x = 0 \qquad (41)$$

The requirement that $\delta J^* = 0$ in spite of arbitrary independent control variations now causes (38) to separate. Using (28), (39), and (40) we have:

$$\bar{u} = P\omega f_u = P\bar{\omega}_{235} \qquad (42)$$

$$\beta = \frac{m_a^2}{2\eta P_D} \int_0^{t_a} (P + P_B)\bar{\omega}_{235} \cdot \bar{\omega}_{235} dt \qquad (43)$$

$$\bar{\omega}_{235}(0) = \frac{\eta(m_D + k_2)}{v_c m_D^2(v_0 + \Delta v_D)} \bar{v}_H \qquad (44)$$

Putting the control (42) into the state equations (27) and (28) gives

$$\dot{x}^T = \left[x_2, \ P\omega_2 - \frac{\mu x_1}{r^3} + \frac{x_3^2}{x_1}, \ P\omega_3 - \frac{x_2 x_3}{x_1}, \ x_5, \ P\omega_5 - \frac{\mu x_4}{r^3} \right] \qquad (45)$$

and expanding (41) with the help of (28) and (42) results in

$$\dot{\omega}_1 = \omega_2 \left[\frac{\mu}{r^3} \left(1 - \frac{3x_1^2}{r^2} \right) + \frac{x_3^2}{x_1^2} \right] - \frac{\omega_3 x_2 x_3}{x_1^2} - \frac{3\mu\omega_5 x_1 x_4}{r^5}$$

$$\dot{\omega}_2 = \frac{\omega_3 x_3}{x_1} - \omega_1$$

$$\dot{\omega}_3 = \frac{1}{x_1} (\omega_3 x_2 - 2\omega_2 x_3) \qquad\qquad\qquad (46)$$

$$\dot{\omega}_4 = \frac{\mu}{r^3} \left[\omega_5 \left(1 - \frac{3x_4^2}{r^2} \right) - \frac{3\omega_2 x_1 x_4}{r^2} \right]$$

$$\dot{\omega}_5 = -\omega_4$$

Finally, we can eliminate the control from (17) to give

$$\frac{1}{m_a} = \frac{1}{m_D} + \frac{1}{2\eta} \int_0^{t_a} P\overline{\omega}_{235} \cdot \overline{\omega}_{235} dt \tag{47}$$

The system (45) and (46) is of 10th order and has 13 boundary conditions (30), (31), and (44), and two path constraints (43) and (47). These are just sufficient to determine the 10 variables x and ω, and the five constants $\overline{v}_H$, P_D, and m_a. The auxiliary parameters Δv_D and m_D are determined from $\overline{v}_H$ using (21) and (22).

9. A Method of Solution

One way to solve the system (45-6) is the "sweep" method[33]. This leads to a 5 x 5 matrix Riccati system, with additional complications arising from the free parameters $\overline{v}_H$ and P_D, and the integral path constraints. It does not look promising.

A conceptually simpler approach is to start at one end, guess the unknown boundary conditions and control parameters, integrate to the other end, see how far off the terminal conditions are, and try again. Six guesses are needed in either case - integrating forward from departure requires that we guess ω_1, $\overline{v}_H$, P_D, and ω_4; while a backward integration from the turn point needs P_D and ω. The former has been chosen, partly because it is mentally easier to follow time's arrow, and partly because the choice of $\overline{v}_H$ at departure is less of a blind guess than ω components at arrival.

With the guess of $\overline{v}_H$ we can determine Δv_D and m_D from (21) and (22), and then $\overline{\omega}_{235}(0)$ from (44). Also, x(0) is found immediately from (31). After computing P(t) from (2) we have everything we need to integrate the system

(45-6). On reaching t_a we can compare the computed $x(t_a)$ with the desired values (30). Also, we can compute m_a from (47), and then a computed value of β from (43). We can compare this with the actual value $\beta = 0.2649$ kg/W. Thus, there are six components of terminal error, corresponding to the six initial guesses.

In order to improve our initial guesses we need to know how all the terminal errors vary with respect to them. If we are close enough to apply linear theory, then this information takes the form of a 6 x 6 matrix of partial derivatives. Perhaps the simplest way to acquire these is to make a nominal run using the initial guesses, and six more runs in each of which one initial guess is varied slightly. By comparing each of these with the nominal, and dividing by the variations we can obtain numerical estimates of the 36 derivatives.

To make this process more explicit, we define an initial guess vector

$$y^T = [\omega_1, \; \bar{v}_H^T, \; P_D, \; \omega_4] \tag{48}$$

and a terminal value vector

$$z^T = [x^T(t_a), \; \beta] \tag{49}$$

Then a typical derivative is determined by

$$\partial_{y_j} z_i = [z_i(y_N + \delta y_j) - z_i(y_N)]/\delta y_j \qquad \text{(no sum on } j) \tag{50}$$

where i and j have the range 1, 2···6; and y_N is the nominal.

Now suppose z_D is the desired terminal vector, comprised of $x(t_a)$ from (30), and $\beta = 0.2649$; and that this can be reached from some specific initial choice y_0. Then in some neighborhood of these we can write:

$$\Delta z \equiv z_D - z = M(y_0 - y) \equiv M\Delta y \tag{51}$$

where

$$M = [M_{ij}] = [\partial_{y_j} z_i] \tag{52}$$

Thus, if δz is sufficiently small, an improved guess y' can be found from:

$$y' = y + M^{-1}\Delta z \tag{53}$$

If the integration is sufficiently accurate, M is nonsingular, and the initial y_0 is close enough, an iteration of this Newton-Raphson procedure should converge fairly rapidly.

The main problem with all this is that for transfer times t_a of several years, z becomes extremely sensitive to small changes in y. This means that the region of convergence around the optimal y is very small, and first rate guessing is required to start the process.

Four methods have been found to deal with the problem. We will call these the "shotgun", the "guided solution", the "2D", and the "second order" methods; and discuss them in this order. First, the shotgun merely means that a series of more or less educated guesses of y are made in the hope that one of the resulting $z - z_D$ vectors will be small enough that the above method will converge. In practice the shotgun has been used only when all else fails.

In the guided solution, suppose we have guessed some y, integrated, and ended up with some z. If the results are at least physically realistic, then we have found the optimal solution to a problem that we might have posed. We can then vary this z toward the desired z, giving us an intermediate goal, and enter the above procedure. After convergence, the process is repeated using the final y as the next starting value, until, in the final step, we jump to the desired z. The region of convergence is greatly extended by this process. We can also guide t_a in this way.

The 2D approach is applicable only for foci lying in the ecliptic; but in those cases it greatly reduces the system complexity and integration time. By solving this special case first, we are then in a position to guide the solution to higher latitudes. Let's look at the reduced problem.

The desired terminal condition (30) is now:

$$x^T(t_a) = [r_T,\ v_T,\ 0,\ 0,\ 0] \qquad (54)$$

We will _guess_ that the solution requires that $\omega_4(0) = \omega_5(0) = 0$. Then from (44) we find that $v_{Hz} = 0$. Putting this into the initial condition (31), and then examining the system (45-6), we can see that part of the solution must be:

$$\omega_4(t) = \omega_5(t) = x_4(t) = x_5(t) = 0 \qquad (55)$$

which is certainly consistent with (54). For such a planar solution we have

$$x_1 = \rho = r \qquad (56)$$

and after dropping the now sterile fourth and fifth elements, the system (45-6) reduces to:

$$\dot{x}^T = \left[x_2, \; P\omega_2 - \frac{\mu}{x_1^2} + \frac{x_3^2}{x_1}, \; P\omega_3 - \frac{x_2 x_3}{x_1} \right] \tag{57}$$

$$= \left[\frac{\omega_2}{x_1^2} \left(x_3^2 - \frac{2\mu}{x_1} \right) - \frac{\omega_3 x_2 x_3}{x_1^2}, \; \frac{\omega_3 x_3}{x_1} - \omega_1, \; \frac{1}{x_1} (\omega_3 x_2 - 2\omega_2 x_3) \right] \tag{58}$$

and the integral conditions (47) and (43) now read:

$$\frac{1}{m_a} = \frac{1}{m_D} + \frac{1}{2\eta} \int_0^{t_a} P(\omega_2^2 + \omega_3^2) dt \tag{59}$$

$$\beta = \frac{m_a^2}{2\eta P_D} \int_0^{t_a} (P + P_B)(\omega_2^2 + \omega_3^2) dt \tag{60}$$

The reduced system is of sixth order and has eight boundary conditions (54), (31), and (44), and two path constraints (59) and (60). These are just sufficient to determine the six variables x and ω, and the four constants P_D, $v_{H\rho}$, $v_{H\phi}$, and m_a. As before, the auxiliary parameters Δv_D and m_D are determined from (21) and (22). Thus the reduced problem is well posed, and any solution to it will be stationary with respect to arbitrary control variations, even in the $\hat{z}$ direction. This shows that our guesses on $\omega_4(0)$ and $\omega_5(0)$ lead to a class of planar solutions; but we have not shown that other solutions do not exist which are consistent with (54).

The final method for improving convergence is the "second order" approach. The relation (51) is really a Taylor series truncated after first order, which is likely to fail for too large a Δy. The truncation error can be reduced, and the feasible Δy increased by extending (51) to second order:

$$z_{Di} = z_i + (\partial_{y_j} z_i)\Delta y_j + \frac{1}{2}(\partial_{y_j y_k} z_i)\Delta y_j \Delta y_k . \tag{61}$$

This can be viewed as a system of quadratic equations in Δy:

$$F_i(\Delta y) \equiv \frac{1}{2} B_{ijk}\Delta y_j \Delta y_k + M_{ij}\Delta y_j + z_i - z_{Di} = 0 \tag{62}$$

in which M is from (52) and

$$B \equiv [B_{ijk}] = [\partial_{y_j} \partial_{y_k} z_i] \tag{63}$$

This is a three dimensional matrix, which is symmetric in its last two indices.

To evaluate B, we can follow the same general method as with M in (50):

$$B_{ijk} = \partial_{y_k} M_{ij} = (\delta y_k)^{-1}[M_{ij}(y_N + \delta y_k) - M_{ij}(y_N)] \qquad \text{(no sums)}$$

$$\tag{64}$$

$$= (\delta y_j \delta y_k)^{-1}[z_i(y_N + \delta y_j + \delta y_k) - z_i(y_N + \delta y_j) - z_i(y_N + \delta y_k) + z_i(y_N)]$$

The set of terms $z_i(y_N + \delta y_j + \delta y_k)$ is new, and requires additional runs parallel to y_N. In two dimensions we need 10 extra runs, for a total of 14; while in three dimensions 21 extras are needed, for a total of 27.

When M and B have been evaluated, it remains to solve (62) for Δy. In two or three dimensions this is a system of 4 or 6 quadratic equations in as many unknowns. No better way to deal with these has been found than the Newton-Raphson method. In this, if we possess some initial guess Δy, an (hopefully) improved Δy is:

IV-23

$$\cdot \; \Delta'y = \Delta y - [\partial_{\Delta y}F]^{-1}F \tag{65}$$

in which, from (62) and the symmetry of B:

$$\partial_{\Delta y_j}F_i = M_{ij} + B_{ijk}\Delta y_k \tag{66}$$

Experience with this second order method has been spotty. It often yields dramatic improvement over the first order method; i.e., $z \rightarrow z_D$; but at other times even the algebraic iteration (65) refuses to converge, which implies that (62) has no real solution. The latter could occur if either the truncation (61) were still too drastic, or if the answer lay outside the region of convergence of the complete Taylor series. In such cases, progress has sometimes been made by looking for an extremal of F with this technique ($F'(\Delta y) = 0$; $F \neq 0$).

Whatever methods have been used, if a solution is reached, we can complete the picture by finding $\phi(t)$. To do this we note from (25) and (26) that

$$\dot{\phi} = v_\phi/\rho = x_3/x_1 \tag{67}$$

from which

$$\phi(t) = \phi(0) + \int_0^t \frac{x_3(\tau)}{x_1(\tau)}\, d\tau \tag{68}$$

The value of $\phi(0)$ is determined from this because of our prior knowledge of $\phi(t_a) = \phi_a$.

A word is needed on the integration methods. First, the integral path constraints need fairly precise evaluation, and the simple trapezoidal rule is not really adequate. A more elaborate method, based on parabolic fitting, is given in Appendix C. It was used in all runs published here. The calculations of $\Delta\phi$ and $\phi(t)$ are done after convergence and are not so critical. Because of this, the two dimensional $\Delta\phi$ calculations were done by the trapezoidal rule, and are thus a little less accurate.

Several schemes were tried out for integration of the differential equations; but none seemed to work as well as a fourth order Runge-Kutta. To put our system into the necessary form, we define an overall state vector:

$$X^T = [x^T, \omega] \qquad (69)$$

and restate our system as:

$$\dot{X} = F(X, t) \qquad (70)$$

Then, if we let h be our integration interval, the Runge-Kutta scheme requires this sequence of steps (see, for instance, Reference 36, page 346):

$$\left.\begin{aligned}
K_1 &= hF(X, t) \\
K_2 &= hF(X + \tfrac{1}{2} K_1, \ t + \tfrac{1}{2} h) \\
K_3 &= hF(X + \tfrac{1}{2} K_2, \ t + \tfrac{1}{2} h) \\
K_4 &= hF(X + K_3, \ t + h) \\
X(t + h) &= \tfrac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4)
\end{aligned}\right\} \qquad (71)$$

In our system, the only explicit time dependence in $F(X, t)$ is the power decay, and this is rather slow. For this reason, and because it was somewhat easier to program, the explicit time dependence in (71) was dropped; but the power was reevaluated before each K_1 and K_3 step. This actually worked very well, as shown by the time step analysis in Appendix D; however, in retrospect, it probably would have been a little more efficient to hew more strictly to (71).

Various possibilities exist for time step control. The simplest is to use a fixed time step h; but, when the action is slow, this is very wasteful of machine time. In interplanetary trajectory work, without strong perturbations, it is common to let $h = Kr^{3/2}$. Here, the ion drive acceleration is often comparable to that of the sun, particularly at large r. Still, the acceleration does fall off with r, so a better choice would be $h = Kr$. By mistake, in most of the three dimensional trajectories, $h = K\rho$ was used. This wasted some computer time at high latitudes; but did not diminish the accuracy. It seems likely that a more efficient choice would take the form $h = K/|\dot{v}|$; but this was not tried.

10. Two Dimensional Program and Results

One final improvement has been made in the 2D system. Since x_1 appears in (57, 8) only as $1/x_1$, we have taken $X_1 = 1/x_1$. When this change is incorporated, the system (70) works out to:

$c \sim 3$

$$\dot{X}_1 = -X_1^2 X_2$$

$$\dot{X}_2 = PX_5 + X_1 X_3^2 - \mu X_1^2$$

$$\dot{X}_3 = PX_6 - X_1 X_2 X_3$$

$$\dot{X}_4 = X_1^2 [X_5(X_3^2 - 2\mu X_1) - X_2 X_3 X_6]$$

$$\dot{X}_5 = X_1 X_3 X_6 - X_4$$

$$\dot{X}_6 = X_1(X_2 X_6 - 2X_3 X_5)$$

(72)

All division has been eliminated and the indexing is a little simpler. Note that y_1 and z_1 are affected by this.

The foregoing structure has been incorporated into the program ION2D and its subprograms FN, RKINT, FNM, PARINT, and VPARINT. The last two are integral evaluation routines, and are covered in Appendix C. The others use a common set of global (APL terminology) variables for their interaction. A translation of these into the notation of the text is given in Table IV-1. The programs themselves are listed below.

```
       V  IUN:0
[1]    FFT*0.36*K1
[2]    H-K1*KF
[3]    HV-P0
[4]    r0<Y
[5]    P*G*(K1l++//F2  3 lv9)*0.5
[6]    nu;U+ K L*  *F l*/0V0-P5*-K19
[7]    L40/ K15/0M0 -RG0*nU/;l0-l0H0  Un0 F15
[8]    P0*YE9]
[9]    )/ FRE.YE9 1.(VE*YE5T/.YE11.KM0*YF2  *1
[10]   Xl*  *  1  , 0*l0*Y
[11]   jjV0,n0
[12]   T*L*0
[13]   0  P *P0**(T*11H:97  PR
[14]   K1*H*l*N
[15]   K2*H*FN  *K1 *V
[16]   P*(P0**T1*/*0.75*H)-P0
[17]   K*  ll*Fn  X*K *-9
[18]   FN-Hxl*N  *K5
[19]   Xn*XH,X(X+*K1*K0 *2*K2/K5)*5
[20]   *(TP*-r*1*H)/w1
```

```
[21]    T,(÷X[1]),X[,3],1000000×X[4 5 6]
[22]  Q1:HV←HV,H
[23]    →L/Q2
[24]    →(T:TA←KKT+H×KT÷X[1])/Q
[25]    H←TA-1
[26]    L←1
[27]    →Q
[28]  Q2:→(T≥TP)/Q3
[29]    T,(÷X[1]),X[,3],1000000×X[4 5 6]
[30]  Q3:IG←÷/[1] XM[5 6 ;]×2
[31]    11←HV PARINT IG
[32]    T2←HV PARINT 10×EV←1,*]×\HV
[33]    MA←(,HD)×(HD×I2)-P%[1]
[34]    Z0←X[,3],I2×HA×2
[35]    ,0
[36]    DPHI←0.5×÷/HV×(11IG→ 1↓IG×÷/[1] XM[] 5 ;]
[37]    Z0[4],DPHI,MA,(MA-BET×P0),NP←,0
[38]    11
[39]  QA:Y←@(4 4 ,p1)←4 4 ,pYE←4 4 ,p0
[40]    RSS←(K11←÷/[1] YE? 3 ; ? 3]*2)*0.5
[41]    DMD←K[2]×K[3]×R%S-K[4]
[42]    KMD←K15×DMD÷RSS×MD×MU←DMD÷K16
[43]    MD←MD0,MD,NMU
[44]    %×XD[5 6],×Y[;3 ; 2 3]×? 2 p←MD),XD[5 6]
[45]    X←N(4p1RE),Y[,2;],(VE←Y[3;]),Y[1;],NX
[46]    NCOL,4
[47]    RKINT
[48]    H←(.←@ 4 4 ,pZ0)÷ 4 4 ,pYE
[49]  QA: [P 0Y,←10×÷0]EH
[50]    Q4:[(Y)×Y0÷0Y-1]
[51]    H←HV[1]
[52]    ,0
[53]  QB:←@(14 4 ,pY)+IM× 14 4 ,pYE
[54]    RSS←(K11←÷/[1] YE? 3 ;INV]*2)*0.5
[55]    DMD←K12-*K15×R%S-K14
[56]    KMD←K15×DMD-RSS×MD×MU←DMD÷K16
[57]    MD←(MU0,MD)[T×3]
[58]    ×←@XD[5 6],Y[? 3 ,GiV]× 2 5 p←MD
[59]    X←N(14p1RE),51,2,7,(VE←Y[3;7),Y[1;],YL1X;1
[60]    NCOL,14
[61]    RKINT
[62]    M←,Y[,4]× 4 4 ,pZ0)÷ 4 4 ,pYF
[63]    R←,5 2 1, 4 4 4 ,pZ0)-B+@U× 2 1 5 @ 4 4 4 ,,X[,4]
[64]    0←(14 4 4 ,pZ[;I×]]÷ 4 4 ,pYE←,XYE
[65]    MC←NY←1
[66]    13←×/1÷T×2K÷Z0 ×2U
[67]    DYF← 4 1 ,p0Y←-FBH
[68]    @5 15,[1]←÷/÷I×2K×(M+0,%'3RE←01,×07)×,×DY
[69]    D←×(D%M,H)←H÷ 00Y←IEH×UR
[70]    ×/←×YF÷10×IDMY)÷Q6
[71]    ←×40? RE←MC÷1)/Q5
[72]  X0DF← 0YME;NY←1 MAFS]
[73]  ×DH← '%LHIV]
[74]  ←,MC,NY,× 1↓15),DY
[75]    →0
        ∇
```

```
    ∇ Z←FN X
[1]    Z←(÷X[2]←X1S),(P×X[5])+(X[1]×X3S←X[3]*2)-MU×X1S←X[1]*2
[2]    Z←Z,((P×X[6])-X[1]×X23),X1S×(X[5]×X3S-TMU×X[1])-X[6]×X23←×/X[2 3]
[3]    Z←Z,((×/X[1 3 6])-X[4]),X[1]×(×/X[2 6])-×/2,X[3 5]
    ∇



    ∇ RKINT
[1]    N←1
[2]    T←0,+\HV
[3]    XKM←(1 2 ,NCOL)ρX[5 6 ;]
[4]  Q:P←(Y[4;]××TI×T[N]+(H←HV[N])÷4)-PB
[5]    K1←H×↑NM X
[6]    K2←H×FNM X+K1÷2
[7]    P←(Y[4;]××TI×T[N]+0.75×H)-PB
[8]    K3←H×FNM X+K2÷2
[9]    K4←H×FNM X+K3
[10]   X←X+(K1+K4+2×K2+K3)÷6
[11]   XKM←XKM,[1] X[5 6 ;]
[12]   →(NP≥N←N+1)/Q
[13]   I1←HV VPARINT IG←+/[2] XKM*2
[14]   I2←HV VPARINT IG×⍉(NCOL,NP)ρEV
[15]   MA←÷(÷MD)+(I2×Y[4;])-PB×I1
[16]   Z←X[,3;],[1] I2×MA*2
    ∇



    ∇ Z←FNM X
[1]    Z←-X[2;]×X1S←X[1;]*2
[2]    Z←Z,[1](P×X[5;])+(X[1;]×X3S←X[3;]*2)-MU×X1S
[3]    Z←Z,[1](P×X[6;])-X[1;]×X23←X[2;]×X[3;]
[4]    Z←Z,[1] X1S×(X[5;]×X3S-TMU×X[1;])-X23×X[6;]
[5]    Z←Z,[1](X[1;]×X[3;]×X[6;])-X[4;]
[6]    Z←(6,NCOL)ρZ,[1] X[1;]×(X[2;]×X[6;])-2×X[3;]×X[5;]
    ∇
```

Table IV-1. ION2D Equivalents

Program Name	Text Notation	Eq.	Program Name	Text Notation	Eq.
B	B	63	M	M	52
BET	$\beta = 0.2649$ kg/watt	1	MA	m_a	17
BR	$M + B\Delta y$	66	MD,MDO	m_D, m_D (nominal)	17
DDY	$\Delta y - \Delta' y$	65	MU	$\mu = 0.1321603\ T_m^3/yr^2$	14,57
DMD	$m_D + k_2$	22,44	N	Index for T, HV	
DPHI	$\phi(t_a) - \phi(0)$	68	NC	Number of iterations	65
DVD	Δv_D	21,44	NCOL	Number of trajectories	
DY	Δy	51	NP	Range of N	
DYM	Δy matrix	65	NY	Index of minimum FS	
EV	$e^{-t/\tau}$	2	P	P	2
F	F	62	PB	$P_B = 100$ watts	2
FS	$F_i F_i$ - Iteration performance	65	PD	P_D	1
H	h	71	RE	$r_e = 0.1495979$ Tm	
HV	h vector	71	RSS	$v_o + \Delta v_D$	21
IG	$\omega_2^2 + \omega_3^2$ vector; also ϕ in ION2D[36]	59,68	T	t	
IM	Index matrix (see below)	73	TA	t_a	
IMV	[2 3 9 10 12]		TI	$-1/\tau = -0.039966/yr$	2

Table IV-1. Continued

Program Name	Text Notation	Eq.	Program Name	Text Notation	Eq.
IRE	$1/r_e = 6.684586/T_m$		TMU	$2\mu = 0.2643206\ T_m^3/yr^2$	58
IX	[1 2 3 1 1 2 3 1 4 5 2 6 3 1]		TP	Output control	
IZ	[5 6 7 8 6 9 10 11 7 10 12 13 8 11 13 14]		VE	$v_e = 0.9399132\ T_m/yr$	
I1	$\int_0^{t_a}(\omega_2^2 + \omega_3^2)dt$	59	X	X (Fixed t)	69
I2	$\int_0^{t_a}(\omega_2^2 + \omega_3^2)e^{-t/\tau}dt$	60	XKM	X (3 Dimensional array)	69
KI1	$2v_0^2 = 0.1210869\ T_m^2/yr^2$	21	XM	X (Matrix - all t)	69
KI2	$k_1 = 29429$ kg	22	X0	X(0)	69
KI3	$1/v_c = 6.941678\ yr/T_m$	22	X1S,X3S	$x_1^2,\ x_3^2$	72
KI4	$v_0 = 0.2460558\ T_m/yr$	21	X23	$x_2 x_3$	72
KI5	$\eta/v_c = 3.470839\ yr/T_m$	44	Y	y	48
KI6	$k_2 = 663.5$ kg	22,44	YE	δy	50,64
KKT	0.36K - Final step parameter	B1	Y0	y (nominal)	48
KMD	Coefficient	44	Z	z	49
KT	K	B1	ZD	z_D	51
K1-4	$K_1 - K_4$	71	ZK	$z_0 - z_D$	62
L	Terminal step switch		Z0	z_0 (nominal z)	49

A series of notes may help in understanding ION2D and its subprograms. Brackets [ ] refer to statement numbers.

[1]   Sets terminal step test. See [24].

[2-4]   Initial values.

[5-7]   Computes Δv_D, m_D, coefficient in (44).

[9]   Constructs X(0).

[10]   Sets up X(t) matrix.

[12]   Sets terminal switch, t = 0.

[13-27]   Integration of main equations (72).

[13, 16]   Power calculations (2).

[14, 15, 17-19] Runge-Kutta structure (71). FN computes F in (70) from (72).

[20, 21]   Output if T ≤ TP.

[22]   Augment h vector.

[23-27]   Final time step.

[28, 29]   Print last step, if not done in [21].

[30-34]   Compute m_a, β, z_0 using (59, 60, 49) Computes vector EV.

[36]   Compute $\Delta\phi$ from (68) using trapezoidal rule.

[37]   Output β, $\Delta\phi$, m_a, J, NP.

[38]   Branch at operator's control.

[39-52]   First order correction loop.

[39]   Matrix of 4 perturbed y vectors.

[40-45]   Repeats [5-9] for 4 parallel trajectories.

[47]   RKINT repeats Runge-Kutta, m_a, β, z calculations for 4 parallel trajectories.

[48]   Computes sensitivity matrix M from (50), (52).

[49]   Computes, displays Δy from (53).

[50]   Computes, displays new y, using operator controlled fraction of Δy.

[51, 52] Restart.

[53-75] Second order correction loop.

[53]   Matrix of 14 perturbed y vectors. See IM note below.

[54-59] Repeats [5-9] for 14 parallel trajectories.

[61]   Same as [47] for 14 trajectories.

[62]   Computes M as in [48].

[63, 64] Computes 4 x 4 x 4 B matrix from (63), (64).

[65-74] Solves system (62) for Δy.

[75]   Branch at operator's control - normally to [50].

It remains to describe the indexing used in the second order loop. In computing B in (64), we need 4 trajectories to get the $z_i(y_N + \delta y_j)$ terms (also needed to compute M), and 10 more for the $z_i(y_N + \delta y_j + \delta y_k)$ terms (the j-k symmetry again). By arranging the latter as the upper triangular elements of a matrix with indices j and k, we get the picture at right: Reading this back out, we can see that the distribution matrix IM in [53] must have the form:

$$
\begin{array}{c}
\quad\quad j \rightarrow \\
k \downarrow
\begin{bmatrix}
5 & 6 & 7 & 8 \\
 & 9 & 10 & 11 \\
 & & 12 & 13 \\
 & & & 14
\end{bmatrix}
\end{array}
$$

$$
IM^T = \begin{bmatrix}
1 & 0 & 0 & 0 & 2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 2 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 2
\end{bmatrix}
\quad\quad (73)
$$

The index vectors IMV, IX, and IZ also arise from this. IMV is used in [54] and [58] to select those trajectories which have a different $\overline{v}_H$ from the nominal. Since this appears only in y_2 and y_3, an examination of (73) shows that IMV = [2 3 9 10 12]. IX is used in [57] to organize the 14 element vector m_D. This is made up from the nominal and the 5 y vectors selected by IMV. Again, reference to (73) shows that its structure is IX = [1 2 3 1 1 2 3 1 4 5 2 6 3 1]. Finally, IZ is used in [64] to index the $z_i(y_N + \delta y_j + \delta y_k)$ terms. It is found by filling in the triangular matrix above to be a symmetrical matrix, and then reading it out serially. Thus, IZ = [5 6 7 8 6 9 10 11 7 10 12 13 8 11 13 14].

ION2D did not always look like this. Initially, with no idea of how to guess y, a greatly simplified version was developed, in which the power decay was ignored, all integral evaluations were done by the trapezoidal rule, and the second order corrector was yet to be developed. A lengthy shotgun finally led to a solution for t_a = 2.5 years. With this as an anchor point, another solution was quickly found at t_a = 2.4 years. Then, using linear extrapolation, the solution was guided in both directions in t_a. Later, this process was refined to a parabolic extrapolation. For this, if y(x) is known at 3 equispaced points y_1, y_2, and y_3, then the parabolic extrapolation is easily shown to be:

$$y_4 = y_1 + 3(y_3 - y_2) \tag{74}$$

At $t_a \approx 5.5$ years a surprising event occurred - convergence could no longer be achieved, even by greatly shortening the extrapolation step. Suspecting that a local conjugate point had been encountered, a longer extrapolation was tried, without success. Lacking an anchor point, the shotgun was used at $t_a \sim 7$ years, again without hitting anything.

Could it be that there are no optimal trajectories for $t_a > 5.5$ years? Consider a trajectory formed by taking a known result for $t_a < 5.5$ years, and inserting an unpowered arc for one or more orbits. Since the power decay has been ignored, this would give the same performance J, but for $t_a > 5.5$ years. Thus, valid trajectories exist for any t_a; but, as unpowered arcs are obviously wasteful, improvements must be possible, and the conclusion that optimal trajectories exist for every t_a seems unavoidable.

This argument suggests that longer times imply more turns around the sun, an effect which has been seen in interplanetary ion drive studies in the past[64]. With this in mind, a new shotgun search was made at $t_a = 8$ years, concentrating in those areas where the integrated $\Delta\phi$ was large. Partly because the second order machinery was then working, the search was successful. An optimal trajectory was found which made about 2.5 turns - a big jump from the results with $t_a < 5.5$ years, where $\Delta\phi$ was less than one turn.

Starting from this anchor point, and using the above extrapolation methods, the results were soon guided up and down in t_a. On the high side, the effort was stopped at $t_a = 10$ years because J was no longer increasing rapidly enough to justify further patience, and because integration of these trajectories was becoming very expensive. On the low side, another blockade was found in that convergence could not be achieved for $t_a < 7.65$ years.

What was going on here? All indications were that optimal trajectories with $\Delta\phi \sim 1.5$ turns should exist for $t_a \sim 6.5$ years; but how would these connect with the known results? The suggestion that we were looking at three distinct, overlapping, classes of trajectories was first made by Dr. Masahiro Kurosaki, and has since been amply verified.

With this insight, the preliminary investigation was deemed complete; but, before searching for the missing trajectories, the program should be

refined. . It was completely rewritten in the present form and debugged on an IBM 5100 computer. The campus IBM 370-168 was then used to redo the previous work for $t_a \leq 5.7$ years (the blockade has shifted upward slightly due to the power decay). The integration step size study in Appendix D was also done at this time. Folowing this, the higher t_a work was repeated, giving results for $t_a \geq 7.8$ years. The most obvious differences were an increase in P_D and a drop in J relative to the earlier runs - the expected effect of including, the power decay.

In searching for an anchor point somewhere in the gap, an improvement was found. in the shotgun approach. Starting from some guessed y, the integration was followed until $r \sim r_T$. On finding this we set $t_a = t$, and computed z. This was repeated with a varied y in an attempt to make $z \to z_D$. The search space is effectively reduced by one dimension in this way. After several hours of search with a 5100 computer, a reasonably close approach was found, and the quest was shifted to the campus 370 machine. There, lock-in . was achieved fairly soon at $t_a = 6.6$ years. Using the above extrapolation techniques, this new class was quickly traced over the range $5.6 \leq t_a \leq 8$ years, before running into new blockades. As expected, there is some overlap on both sides.

All these results are summarized in Figures IV-2, 3. The division of the trajectories into classes, as suggested by Kurosaki, is strikingly evi-dent - most plainly in $\Delta\phi$, but visible to some extent in all the other curves In fact, we will find it convenient to label these classes as I, II, or III, depending on whether they are centered about 0.5, 1.5 or 2.5 turns respective Of course, further classes almost surely exist for $t_a > 10$ years.
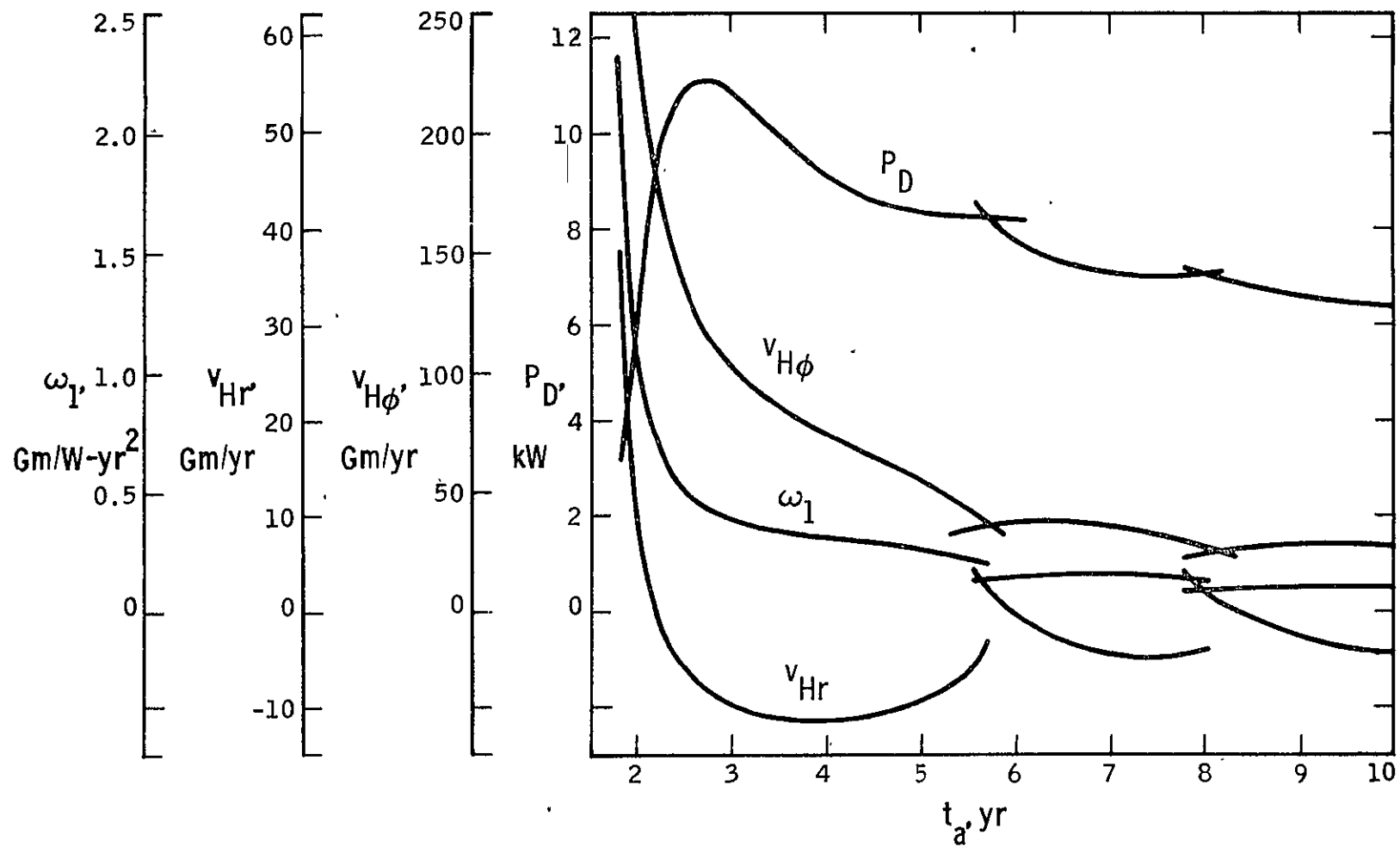
Figure IV-2. Two-Dimensional Trajectories--Initial Conditions
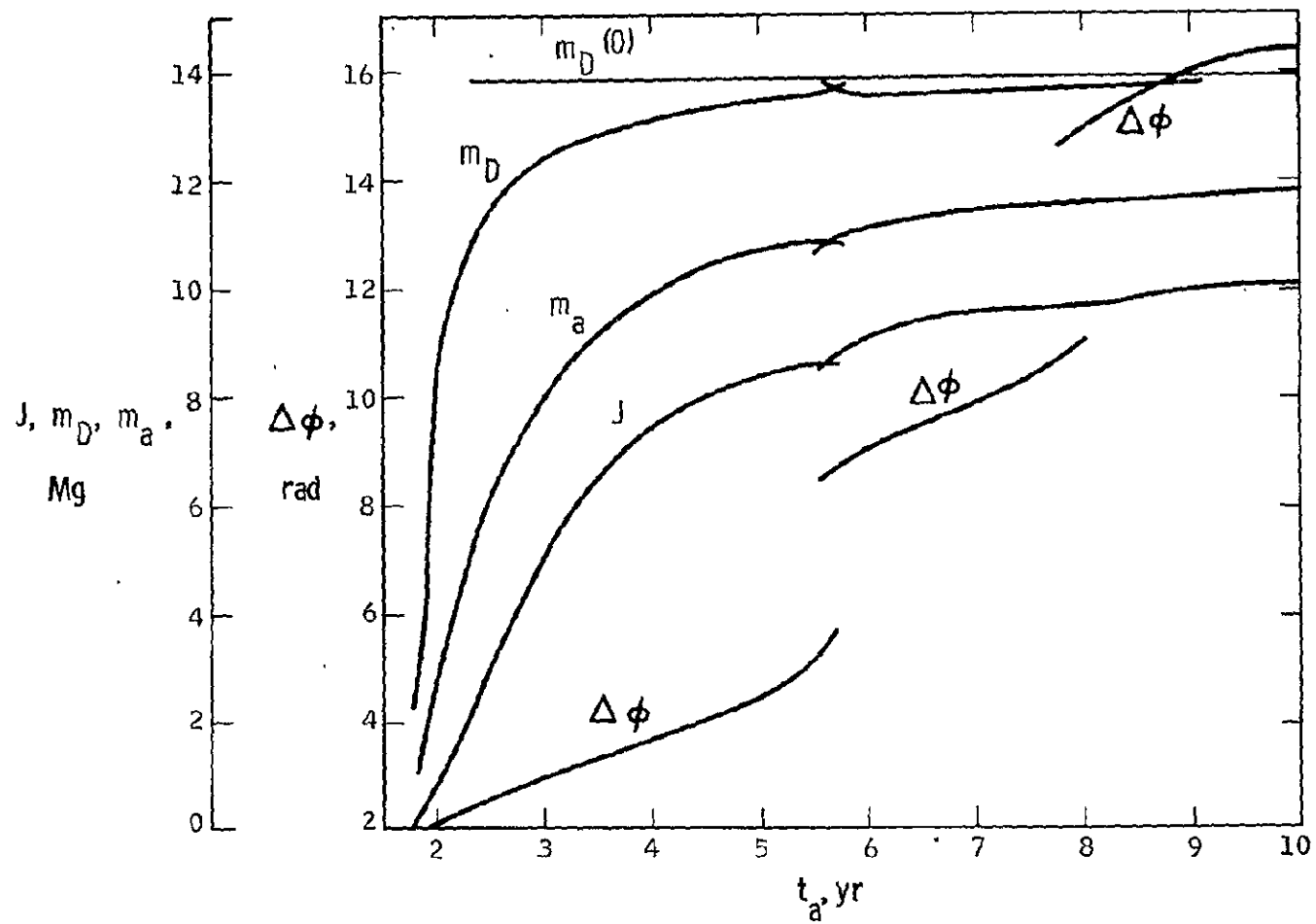
Figure IV-3. Two-Dimensional Trajectories--Performance

There are many interesting features on these graphs. Of greatest practical interest is the performance J. This quickly climbs to ~8 Mg at t_a = 4.5 years, and then much more slowly to ~10 Mg at t_a = 9.5 years. After allowing for the free power supply, this is an order of magntiude better than the optimal chemical payload in the last chapter.

It is of interest to see how close we are to the theoretical limits. Within the present assumptions, the highest possible departure mass m_D comes when $\overline{v}_H$ = 0. From (21)

$$\Delta v_D(0) = (\sqrt{2} - 1)v_0 = 0.10192 \text{ Tm/yr}$$

and putting this into (22) yields $m_D(0)$ = 13.841 Mg. A glance at Figure IV-3 shows that m_D hovers about this limit when t_a > 5 years. Also slowly approaching this limit is the mass at arrival, m_a. At t_a = 10 years, m_a = 11.764 Mg, so it hasn't far to go. This also shows that the ion propellant spent, $m_D - m_a$, will be in the range 2-3 Mg for likely missions, although it is much higher for short times.

Returning to J, it is obvious that as $t_a \to \infty$, $J \to m_a \to m_D(0)$. Thus at t_a = 5 years, 60% of the ultimate performance has already been reached, a figure which climbs to 84% at t_a = 9 years. This shows that there is little practical reason to explore the region t_a > 10 years.

The mass of the power supply, m_a - J, can also be read from these curves, although it is somewhat easier to see from P_D in Figure IV-2. Reasonable trajectories have P_D < 8.4 kW, so that $m_a - J = \beta P_D$ < 2.22 Mg. Although this power is 15-20 times higher than Voyager, it is probably feasible. It is interesting that the estimate (7), $m_D - m_a = \beta P_D$, based on very crude reasoning, is not far off.

Lastly we might look at the components of $\bar{v}_H$ in Figure IV-2. The general trend is for $|\bar{v}_H| \to 0$, although as seen from m_D this is not monotonic. It might be wondered why we don't get $m_D \to m_D(0)$ more rapidly, since the ion drive is so much more efficient than the first stage. The reason is that from (21):

$$\bar{v}_H \cdot d\bar{v}_H = (v_0 + \Delta v_D)d\Delta v_D \tag{75}$$

so that as $\bar{v}_H \to 0$, $d\bar{v}/d\Delta v_D \to \infty$. Thus for sufficiently small $\bar{v}_H$, the first stage is actually more efficient; and for finite t_a, we will always have $|\bar{v}_H| > 0$. Another interesting point is that the late members of each class all have $v_{Hr} < 0$, so that many trajectories begin by heading slightly sunward.

To plot these trajectories we need the longitude, which has earlier been eliminated from the state. However, after a run of ION2D, we can find the ϕ vector by integrating (68), by the trapezoidal rule. This requires the APL statement

PHI←0,(90÷o1)x+\HVx(1↓IG)-¯1↓IG←x/X[1 3]

More precisely, this generates $\phi(t) - \phi(0)$ in degrees, at each t used in $0 \le t \le t_a$.

With this technique, the trajectories for $t_a = 5$, 7, and 9 years have been plotted, and are shown in Figures IV-4-6, illustrating one trajectory in each class. Perhaps the most interesting feature is the similarity to pumping a swing, starting from rest. By examining print outs of the state (not shown), one sees from ω_2 and ω_3 that the thrust is primarily tangential,
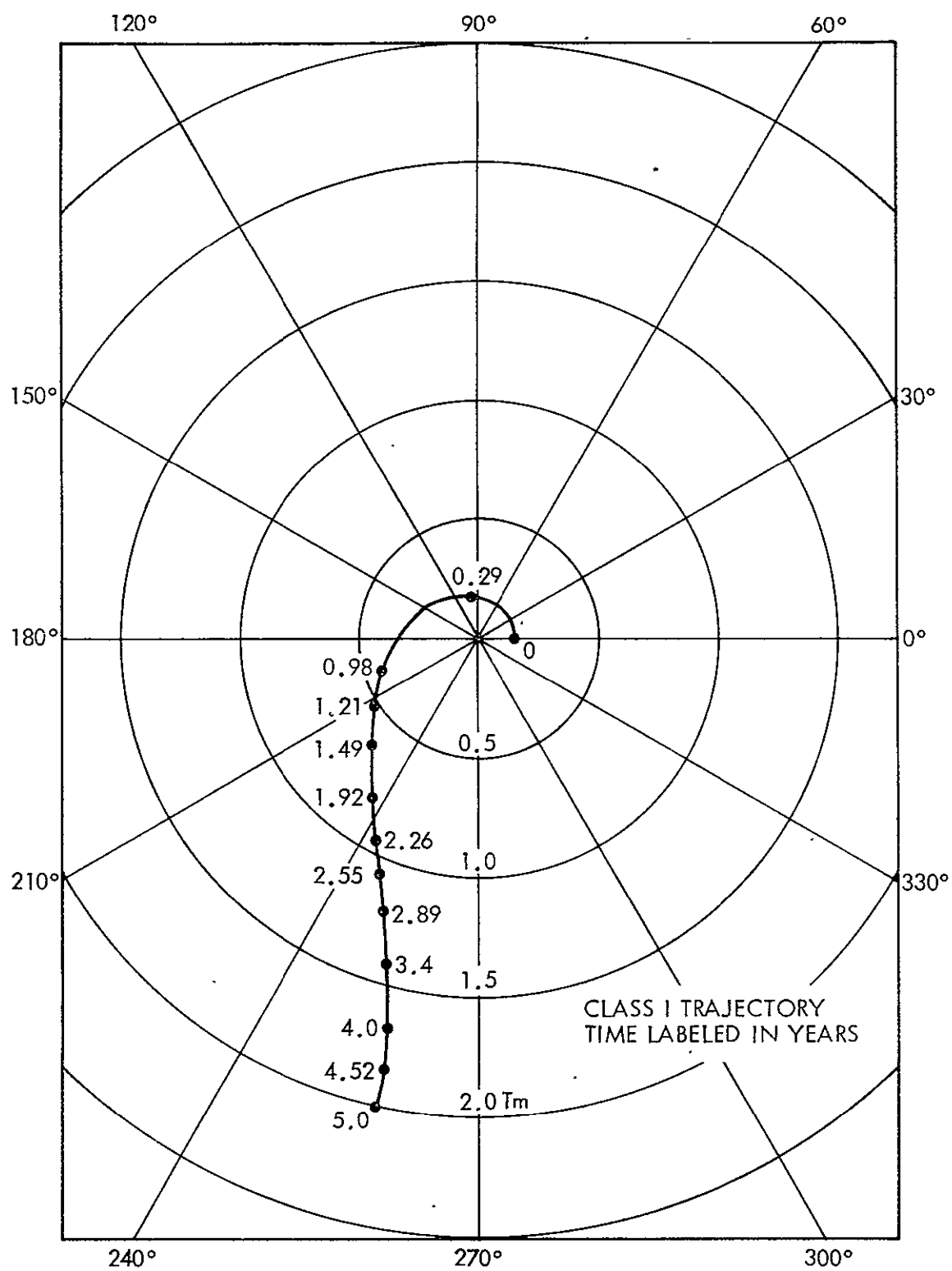
Figure IV-4. Two-Dimensional Trajectory for t_a = 5 Years
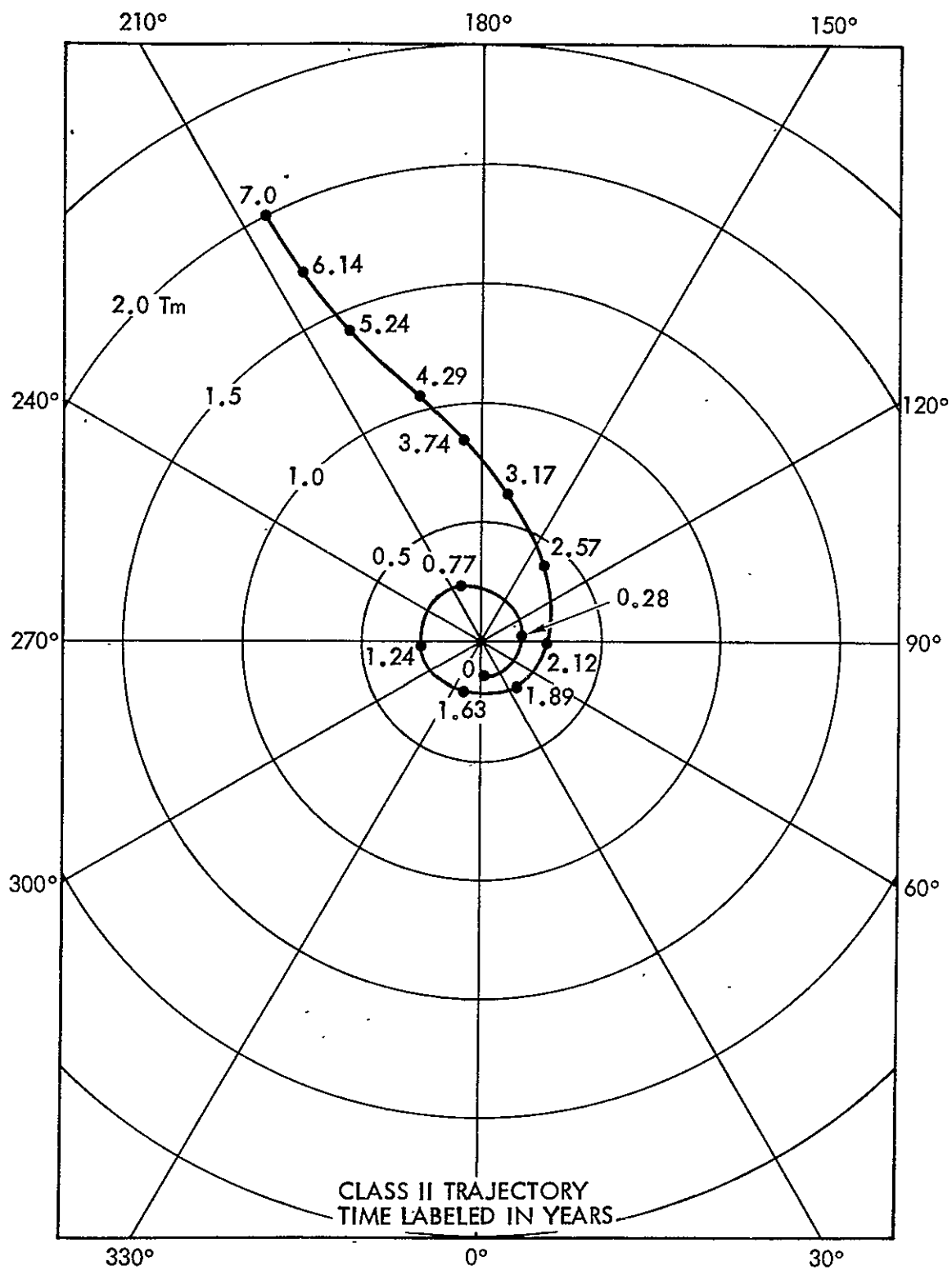
Figure IV-5. Dimensional Trajectory for t_a = 7 Years
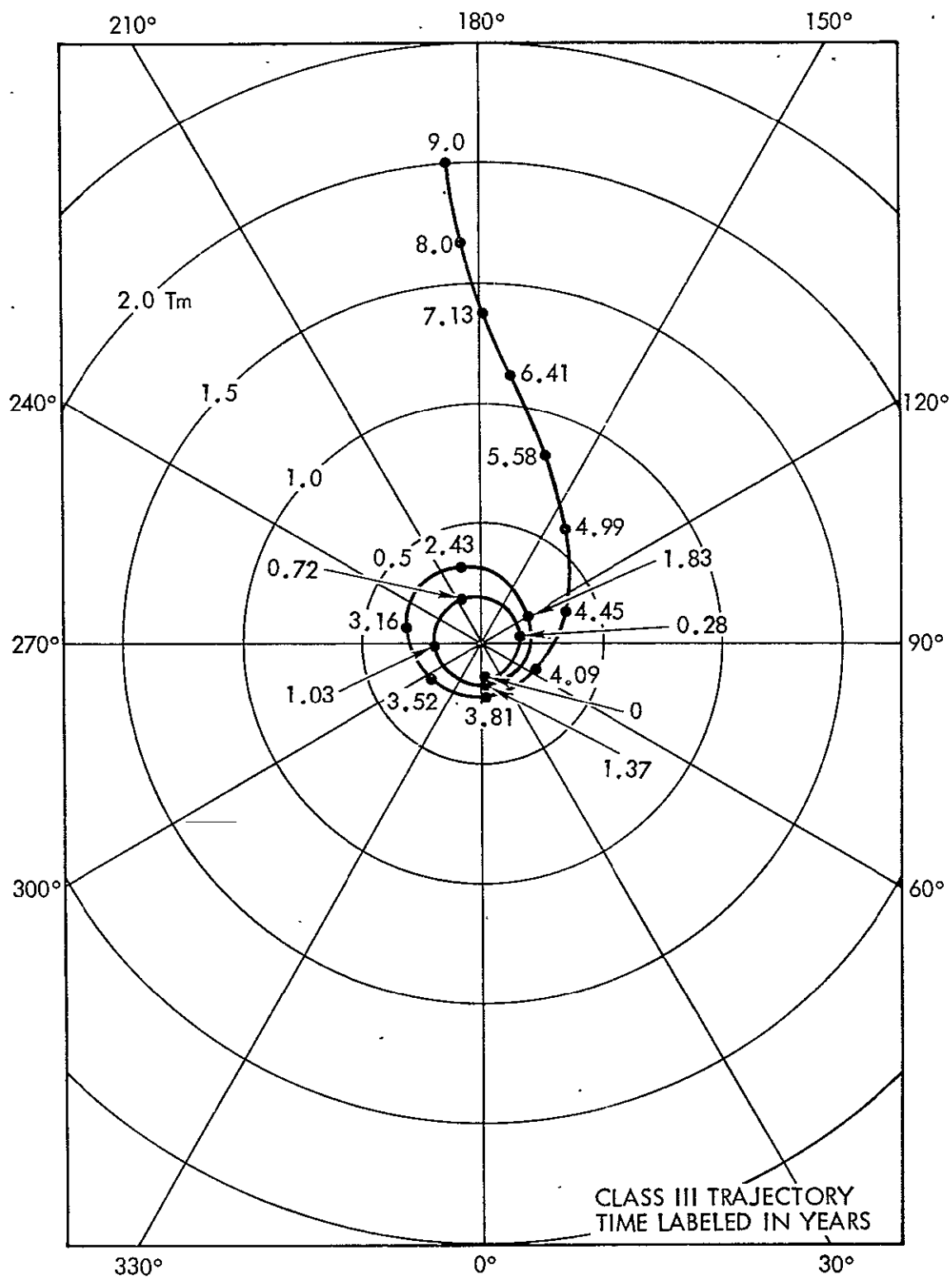
Figure IV-6. Two-Dimensional Trajectory for t_a = 9 Years

and that it is much stronger near the perihelia. The reason for this is that there is a large energy differential between the initial and terminal conditions, and this is most efficiently made up when $|\bar{v}|$ is largest.

11. Three Dimensional Program and Results

In returning to the full 3D system (45, 6), it appears that the adoption of $1/\rho$ as a state variable is not nearly as useful; and that change is now dropped. The full system in the form (70) is now:

$$
\left\{
\begin{aligned}
\dot{X}_1 &= X_2 \\
\dot{X}_2 &= PX_7 + X_1^{-1}X_3^2 - \mu q^3 X_1 \\
\dot{X}_3 &= PX_8 - X_1^{-1}X_2X_3 \\
\dot{X}_4 &= X_5 \\
\dot{X}_5 &= PX_{10} - \mu q^3 X_4 \\
\dot{X}_6 &= X_7[\mu q^3(1 - 3q^2X_1^2) + X_1^{-2}X_3^2] - X_1^{-2}X_2X_3X_8 - 3\mu q^5 X_1 X_4 X_{10} \\
\dot{X}_7 &= X_1^{-1}X_3X_8 - X_6 \\
\dot{X}_8 &= X_1^{-1}(X_2X_8 - 2X_3X_7) \\
\dot{X}_9 &= \mu q^3[X_{10}(1 - 3q^2X_4^2) - 3q^2X_1X_4X_7] \\
\dot{X}_{10} &= -X_9
\end{aligned}
\right.
\tag{76}
$$

where:

$$
q = 1/r = (X_1^2 + X_4^2)^{-1/2}
\tag{77}
$$

All this forms the structure of the program ION3D, and its subprograms

F3, RK3, and FM3. Of these, F3 and FM3 are the state equations (76),

corresponding to FN and FNM in ION2D. RK3 is similar to its 2D counterpart

RKINT, but with minor modifications. The integral evaluation routines

PARINT and VPARINT are unchanged. The variable names have changed quite a

bit, so a new translation is given in Table IV-2. The programs are listed

below.

Table IV-2. ION3D Equivalents

Program Name	Text Notation	Eq.	Program Name	Text Notation	Eq.
CT	$c\theta_a$	30	J7	[1 2 4 6]	
DPH	$\phi(t_a) - \phi(0)$	68	J8	[7 2 1 8 6 3 4 9 5]	
DR	$\pi/180$		MIL	10^6	
IG	$\bar{\omega}_{235} \cdot \bar{\omega}_{235}$ vector	43,47	NR	Number of trajectories	
IM3	Index matrix. See below		RVT	$r_T, v_T = 2, 0.3155693$	30
IVX	[1 8 3 2 9 7 4 5 10 6]		ST	$s\theta_a$	30
I1	$\int_0^{t_a} \bar{\omega}_{235} \cdot \bar{\omega}_{235} dt$	43	SX,SXX	6 6 ; 6 6 6	
I2	$\int_0^{t_a} \bar{\omega}_{235} \cdot \bar{\omega}_{235} e^{-t/\tau} dt$	47	THA	θ_a in degrees	30
J1	[1 7 12]	76	VB	[2 3 4 13 14 15 18 19 22]	
J2	[1 4 10 11 12]	76	VD	Index vector. See below	
J3	[2 8 14]	76	VI	[2 3 4] Selects $\bar{v}_H$ from y	
J4	[4 4 12]	76	YE	$\delta y = [10^{-8}\ 2 \times 10^{-8}\ 10^{-7}\ 10^{-7}\ .1\ 10^{-7}]$	50,64
J5	[1 4 7 12]	76	ZB	Index vector. See below.	

Other labels have the same meanings as in ION2D, and are given in Table IV-1. Vectors and matrices are usually larger, however. In alphameric order these are: B, BET, BR, DDY, DMD, DVD, DY, DYM, EV, F, FS, H, HV, KI1-6, KKT, KMD, KT, K1-4, L, M, MA, MD, MDO, N, NC, NP, NY, P, PB, PD, RE, RSS, T, TA, TI, TP, VE, X, XM, XO, Y, YE, YO, Z, ZD, ZK, ZO.

ION3D is structurally similar to ION2D, but a number of improvements have been made. The notes given here mainly address these differences. Again, [ ] indicates statement numbers.

[1, 2]   Computes z_D from (30) and (44).

[3-12]   Equivalent to ION2D [1-12].

[13-27]   Equivalent to ION2D [13-27].

[28-33]   Equivalent to ION2D [30-37].

[34]   Prints out terminal errors.

[35]   Branch at operator's control.

[36-49]   First order loop. Equivalent to ION2D [39-52].

[50-72]   Second order loop. Equivalent to ION2D [53-75].

Indexing in ION3D is more complicated than before, but not fundamentally different. First, the vectors J1-5 in Table IV-2 are for indexing the calculations (76) in F3 and FM3. If the reader is only slightly familiar with APL, it will be instructive to work this out. The index vectors IVX, J7, and J8 are all used in constructing the X(0) matrices in the first and second order loops. They are purely functions of the ordering of the X and Y vectors, and the order in which these elements are assembled in [41-2] and [55-6].

The scheme for computing B exactly parallels that in ION2D. This time the upper triangular matrix is as shown at right. Reading this one out yields the indexing matrix IM3:

$$
\begin{array}{c}
j \rightarrow \\
\begin{array}{c} k \\ \downarrow \end{array}
\begin{bmatrix}
7 & 8 & 9 & 10 & 11 & 12 \\
  & 13 & 14 & 15 & 16 & 17 \\
  &   & 18 & 19 & 20 & 21 \\
  &   &   & 22 & 23 & 24 \\
  &   &   &   & 25 & 26 \\
  &   &   &   &   & 27
\end{bmatrix}
\end{array}
$$

$$IM3^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 2 \end{bmatrix} \tag{78}$$

Like IMV in ION2D, VB is used to select out those trajectories with non-nominal $\bar{v}_H$. The entry in Table IV-2 is easily read out from (78). VD corresponds to IX in ION2D, and is here used to organize the 27 element vector m_D from the nominal and the 9 trajectories selected by VB. This is readily seen to be:

$$VD = [1 \ 2 \ 3 \ 4 \ 1 \ 1 \ 1 \ 2 \ 3 \ 4 \ 1 \ 1 \ 5 \ 6 \ 7 \ 2 \ 2 \ 8 \ 9 \ 3 \ 3 \ 10 \ 4 \ 4 \ 1 \ 1 \ 1]$$

Finally, ZB, like its counterpart IZ in ION2D, is made up by filling out the upper triangular matrix above into a symmetric matrix, and reading it out serially. This leads to the 36 element vector:

$$ZB = [7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 8 \ 13 \ 14 \ 15 \ 16 \ 17 \ 9 \ 14 \ 18 \ 19 \ 20 \ 21 \ 10 \ 15$$
$$19 \ 22 \ 23 \ 24 \ 11 \ 16 \ 20 \ 23 \ 25 \ 26 \ 12 \ 17 \ 21 \ 24 \ 26 \ 27]$$

Once ION3D had been debugged, the question was, how to use it? Although the many improvements caused it to run nearly as fast as ION2D, an exploration as broad as in two dimensions was out of the question. Eventually it was decided that representative results might be obtained by taking a few values of t_a and calculating optimal trajectories over the whole range of θ_a. This could be done by starting from the known results for $\theta_a = 0$, and guiding the solution toward increasing θ_a. The performance is obviously an even function of θ_a; so there is no need to run $\theta_a < 0$.

The choice of t_a = 5, 7, and 9 years seemed good because it gave a good example of each class of trajectory. In each case, once the process had been started at θ_a = 0, it was found that with only the first order corrector, and quadratic extrapolation, convergence was quickly achieved with 2 deg jumps in θ_a. Much of this work was run on the IBM 370, but it was later found that it would fit on the IBM 5100 if the second order corrector was removed, and provisions were added to erase unused arrays. After the time step calculation was revised, it was found that with K = 0.3, all of the 5 and 7 year runs would fit. Typically, a single 7 year run-correction-rerun sequence would take about an hour on the 5100.

The results were surprising in several ways. First, with t_a = 5 years, the complete sequence up to θ_a = 88 deg was completed without a hitch. (No cases with θ_a = 90 deg were attempted, because the ecliptic pole can't be approached with cylindrical coordinates and $1/\rho$ appears in the state equations.) However, when t_a = 7 years, a blockade was found after θ_a = 52 deg. As with the 2D class limits, first and second order methods, tightening the θ_a jump, and going well past 52 deg all failed to give any result. Temporarily abandoning this, the 9 year series was begun, and completed through θ_a = 88 deg, a considerable surprise.

An indication of what was happening came from the $\Delta\phi$ results. For t_a = 5 and 9 years, this did not vary much from end to end, so these are Class I and III respectively for all θ_a. For 7 years however, $\Delta\phi$ began to turn down sharply near the blockade, although staying within Class II. This suggested a class limit, and that the remaining trajectories would be in Class I.

How to find them? A blind 3D shotgun was not at all attractive; but if the class hypothesis was right, there would be no blockade between the desired

7 years paths and the now existing 5 year results. Thus it was reasonable to try guiding a solution in t_a. The starting point selected was $t_a' =$ 5 years, θ_a = 56 deg. With quadratic extrapolation and 0.1 year jumps, the t_a = 7 year class I trajectory was found in about two days on the 5100. Then working backward in 1 deg steps the expected class limit was found at θ_a = 44 deg, a considerable overlap with the class II results. Going forward in 2 deg steps, θ_a = 88 deg was quickly reached, all of this requiring another two days on the 5100.

All these longitudinal and class features are illustrated in Figure IV-7, in which the bridge to the 7 year class I series is shown as a vertical line at θ_a = 56 deg. Of course what is really interesting here is the performance. Figure IV-8 shows J for all these cases. It had been expected that J would fall off with increasing θ_a, but not this slowly. Indeed, these results are startling compared to the total inability to reach high latitudes by two impulse transfer (see Figure III-9). Other trajectory parameters of some interest are y, Δv_D, m_D, and m_a. All these are given in detail in Tables IV-3-5. Finally, to see what these trajectories look like, the cases θ_a = 40 deg, 80 deg for each t_a are plotted in Figures IV-9-14. These are cylindrical projections onto the ecliptic, with t and z tabled on each plot to assist in visualizing the path. The portions with z < 0 are dashed as a further aid.
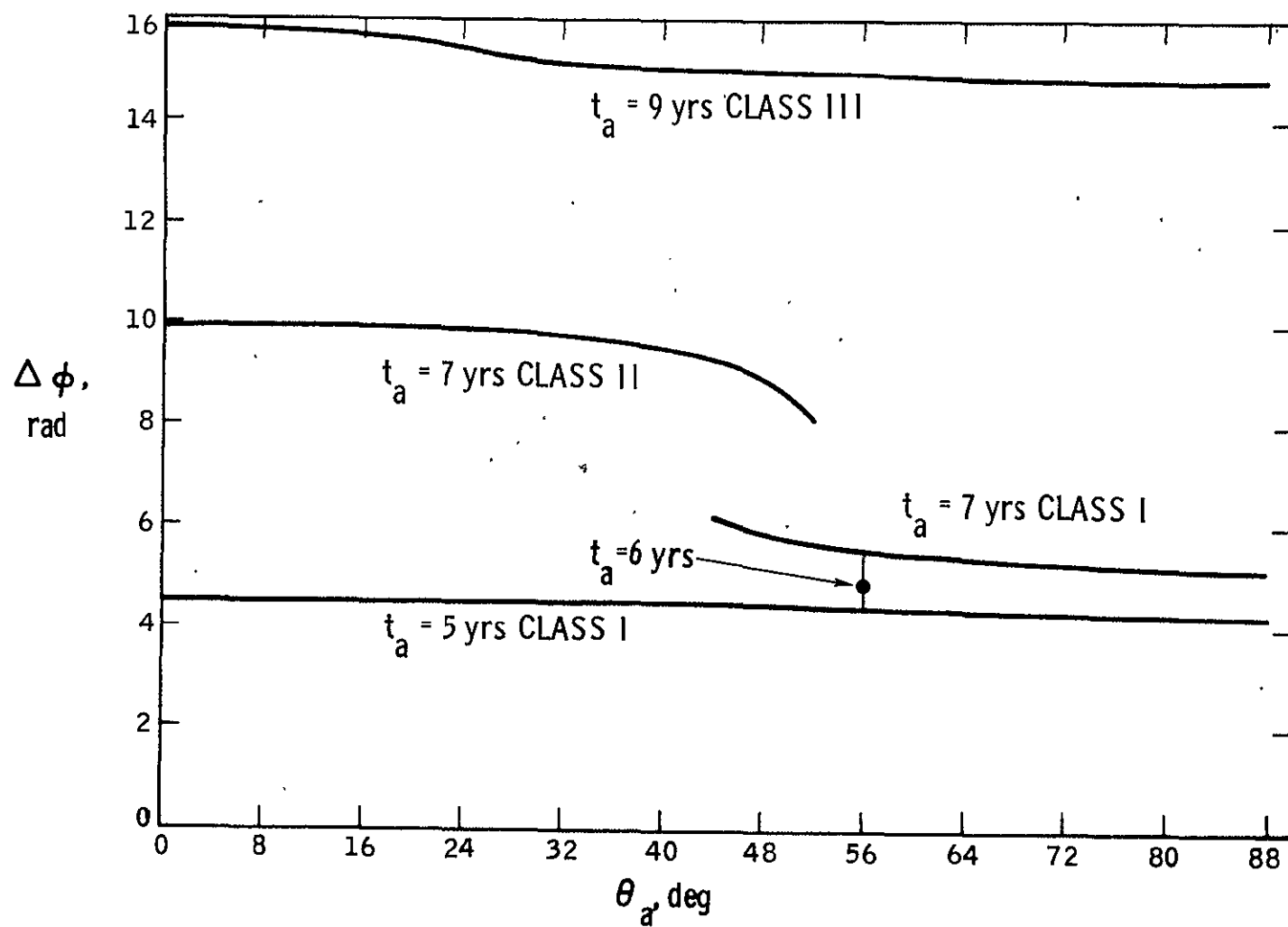
Figure IV-7. Longitude Difference in Three-Dimensional Trajectories
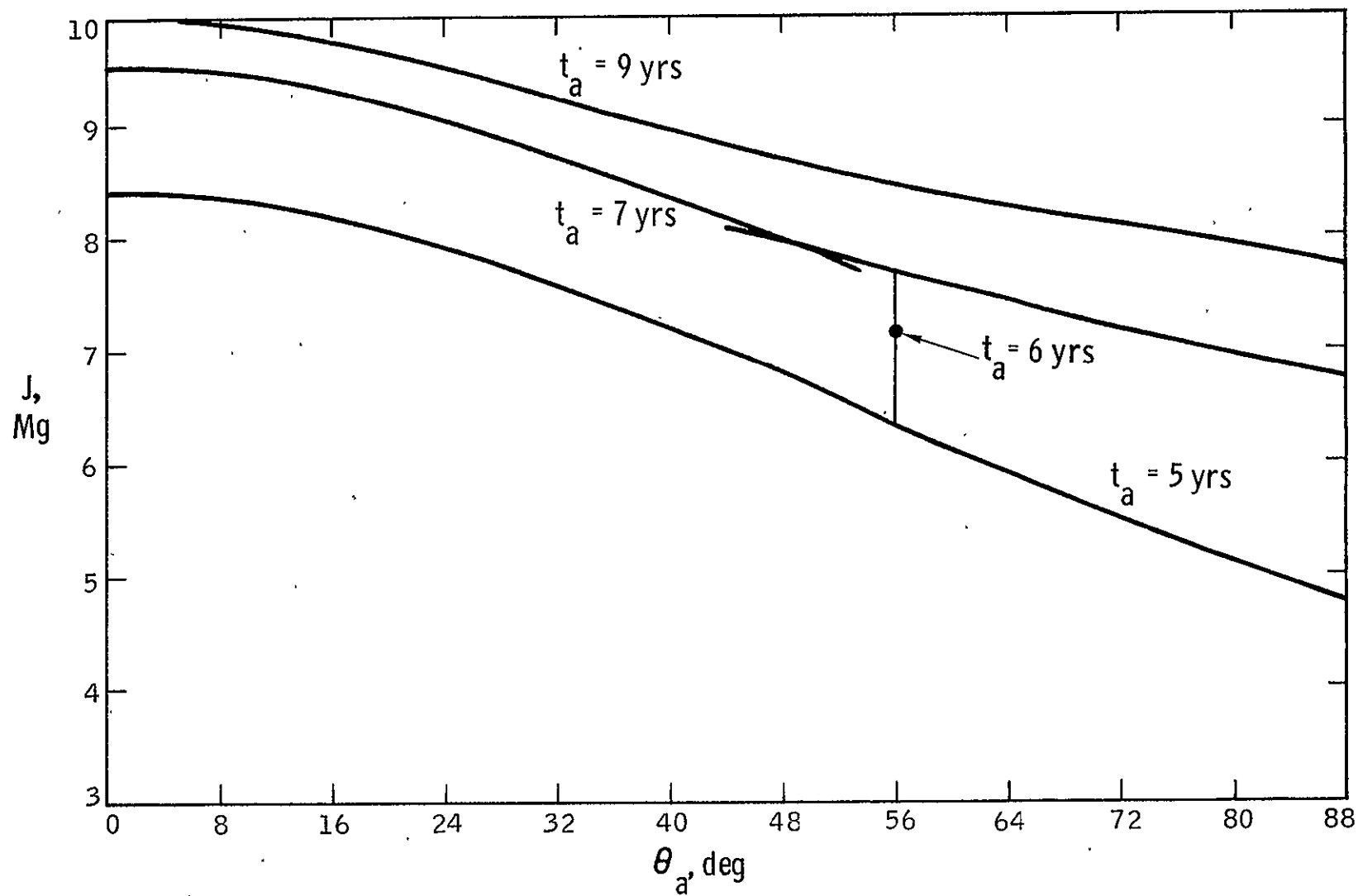
Figure IV-8. Performance of Three-Dimensional Trajectories

Table IV-3. 3D Performance t_a = 5 Years

θ_a deg	J kg	m_a kg	m_D kg	P_D W	$-v_{H\rho}$ Gm/yr	$v_{H\phi}$ Gm/yr	$-v_{Hz}$ Gm/yr	$\Delta\phi$ rad
0	8414	10635	13397	8382	9.201	55.25	0	4.4956
2	8411	10632	13397	8387	9.198	55.227	1.29	4.4957
4	8399	10626	13398	8404	9.191	55.158	2.575	4.4961
6	8381	10614	13399	8431	9.178	55.045	3.847	4.4968
8	8355	10598	13399	8468	9.162	54.893	5.102	4.4976
10	8321	10577	13400	8516	9.141	54.707	6.334	4.4985
12	8281	10552	13401	8573	9.118	54.493	7.539	8.4994
14	8234	10523	13402	8639	9.091	54.257	8.711	4.5002
16	8181	10489	13403	8712	9.062	54.007	9.848	4.5008
18	8122	10452	13404	8794	9.032	54.748	10.946	4.5011
20	8058	10411	13404	8882	9.0	53.488	12.003	4.5011
22	7988	10366	13405	8976	8.968	53.232	13.016	4.5007
24	7913	10317	13405	9076	8.935	52.986	13.984	4.4998
26	7833	10266	13405	9180	8.901	52.753	14.906	4.4985
28	7751	10211	13404	9288	8.868	52.539	15.781	4.4965
30	7664	_10153	13404	9399	8.835	52.344	16.661	4.4941
32	7573	10093	13402	9513	8.801	52.175	17.392	4.4911
34	7480	10030	13401	9628	8.767	52.031	18.129	4.4876
36	7384	9965	13399	9745	8.733	51.916	18.82	4.4835
38	7286	9898	13397	9863	8.698	51.828	19.467	4.4789
40	7186	9830	13394	9981	8.662	51.771	20.07	4.4737
42	7084	9759	13392	10099	8.625	51.746	20.632	4.4681
44	6980	9687	13389	10216	8.587	51.751	21.152	4.462
46	6876	9613	13386	10333	8.546	51.788	21.632	4.4555
48	6771	9539	13382	10447	8.503	51.857	22.073	4.4485
50	6666	9463	13378	10560	8.458	51.956	22.477	4.4412
52	6559	9386	13374	10671	8.41	52.087	22.845	4.4334
54	6453	9309	13370	10780	8.358	52.249	23.177	4.4253
56	6347	9231	13365	10885	8.303	52.44	23.474	4.4169
58	6241	9152	13360	10988	8.244	52.661	23.739	4.4082

Table IV-3.   3D Performance t_a = 5 Years Concluded

θ_a deg	J kg	m_a kg	m_D kg	P_D W	$-v_{H\rho}$ Gm/yr	$v_{H\phi}$ Gm/yr	$-v_{Hz}$ Gm/yr	$\Delta\phi$ rad
60	6136	9073	13355	11088	8.181	52.911	23.972	4.3992
62	6031	8994	13350	11185	8.114	53.189	24.174	4.3899
64	5927	8914	13345	11278	8.042	53.493	24.345	4.3804
66	5823	8835	13339	11368	7.965	53.824	24.488	4.3707
68	5721	8755	13333	11455	7.883	54.18	24.602	4.3607
70	5619	8676	13327	11537	7.796	54.561	24.689	4.3507
72	5519	8596	13321	11616	7.703	54.964	24.75	4.3404
74	5420	8517	13314	11691	7.606	55.389	24.785	4.3299
76	5322	8438	13308	11763	7.503	55.836	24.796	4.3193
78	5225	8359	13301	11830	7.394	56.302	24.783	4.3086
80	5130	8281	13294	11894	7.281	56.787	24.746	4.2978
82	5037	8203	13286	11954	7.162	57.29	24.688	4.2869
84	4945	8126	13279	12011	7.038	57.809	24.607	4.2759
86	4854	8049	13271	12063	6.909	58.343	24.506	4.2649
88	4765	7974	13264	12112	6.775	58.892	24.385	4.2539

Table IV-4.   3D Performance t_a = 7 Years
NOTE: Class Break for θ_a > 48 deg

θ_a deg	J kg	m_a kg	m_D kg	P_D W	$-v_{H\rho}$ Gm/yr	$v_{H\phi}$ Gm/yr	$-v_{Hz}$ Gm/yr	$\Delta\phi$ rad
0	9525	11417	13643	7143	4.532	36.897	0	9.8962
2	9521	11415	13644	7149	4.525	36.863	0.53	9.8961
4	9510	11409	13645	7168	4.504	36.763	1.055	9.8957
6	9491	11398	13646	7199	4.47	36.598	1.572	9.895
8	9465	11383	13648	7242	4.421	36.374	2.075	9.8938
10	9432	11364	13651	7296	4.357	36.096	2.56	9.8921
12	9392	11342	13654	7361	4.276	35.772	3.021	9.8896
14	9345	11315	13657	7436	4.178	35.408	3.453	9.8861

θ_a deg	J kg	m_a kg	m_D kg	P_D W	$-v_{H\rho}$ Gm/yr	$v_{H\phi}$ Gm/yr	$-v_{Hz}$ Gm/yr	$\Delta\phi$ rad
16	9292	11285	13661	7521	4.061	35.014	3.85	9.8813
18	9234	11251	13665	7614	3.922	34.6	4.203	9.875
20	9170	11214	13669	7714	3.76	34.173	4.506	9.8668
22	9102	11174	13673	7821	3.571	33.743	4.749	9.8563
24	9029	11131	13677	7934	3.352	33.319	4.922	9.843
26	8952	11085	13681	8052	3.099	32.909	5.015	9.8264
28	8872	11037	13685	8173	2.806	32.52	5.012	9.8058
30	8788	10986	13689	8298	2.47	32.16	4.898	9.7803
32	8702	10934	13692	8426	2.083	31.832	4.652	9.749
34	8614	10880	13696	8555	1.638	31.538	4.249	9.7105
36	8524	10825	13699	8685	1.126	31.227	3.651	9.6628
38	8433	10768	13702	8815	0.532	31.04	2.81	9.6032
40	8341	10710	13705	8945	-0.158	30.809	1.648	9.5272
42	8248	10652	13707	9074	-0.968	30.543	.052	9.4281
44	8156	10593	13710	9201	-1.911	30.115	-2.136	9.2952
46	8064	10534	13712	9326	-2.962	29.466	-5.041	9.1152
48	7973	10476	13714	9449	-3.974	28.203	-8.519	8.8803
50	7907	10393	13612	9386	5.948	39.26	5.058	5.8361
52	7844	10347	13600	9451	6.276	40.202	5.892	5.7604
54	7780	10301	13589	9517	6.556	41.009	6.6	5.6957
56	7717	10256	13579	9584	6.801	41.717	7.215	5.639
58	7654	10210	13570	9652	7.019	42.351	7.758	5.5883
60	7590	10165	13561	9720	7.215	42.926	8.243	5.5425
62	7527	10120	13554	9788	7.394	43.454	8.678	5.5006
64	7464	10075	13546	9855	7.558	43.944	9.07	5.4627
66	7401	10030	13539	9924	7.709	44.4	9.425	5.4267
68	7338	9985	13533	9992	7.849	44.828	9.746	5.3932
70	7275	9940	13527	10059	7.98	45.233	10.037	5.3617
72	7213	9895	13521	10126	8.102	45.618	10.3	5.332

Table IV-4. 3D Performance t_a = 7 Years Concluded

NOTE: Class Break for θ_a > 48 deg

θ_a deg	J kg	m_a kg	m_D kg	P_D W	$-v_{H\rho}$ Gm/yr	$v_{H\phi}$ Gm/yr	$-v_{Hz}$ Gm/yr	$\Delta\phi$ rad
74	7151	9850	13515	10192	8.216	45.986	10.537	5.3039
76	7079	9806	13509	10257	8.324	46.34	10.75	5.2772
78	7027	9762	13504	10322	8.425	46.681	10.94	5.2517
80	6966	9718	13499	10386	8.52	47.01	11.11	5.2273
82	6909	9674	13494	10448	8.61	47.331	11.26	5.2038
84	6846	9630	13489	10510	8.695	47.643	11.391	5.1811
86	6787	9587	13485	10570	8.776	47.948	11.505	5.158
88	6729	9545	13480	10629	8.852	48.251	11.602	5.1312

Table IV-5. 3D Performance t_a = 9 Years

θ_a deg	J kg	m_a kg	m_D kg	P_D W	$-v_{H\rho}$ Gm/yr	$v_{H\phi}$ Gm/yr	v_{Hz} Gm/yr	$\Delta\phi$ rad
0	9952	11706	13727	6623	2.584	28.115	0	15.766
2	9948	11704	13727	6628	2.577	28.103	-.056	15.764
4	9938	11698	13727	6644	2.556	28.066	-0.101	15.759
6	9922	11689	13728	6671	2.521	28.008	-0.124	15.75
8	9898	11675	13728	6709	2.47	27.933	-0.113	15.737
10	9868	11658	13729	6756	2.409	27.846	-.054	15.719
12	9833	11637	13730	6813	2.317	27.775	.067	15.697
14	9791	11613	13731	6878	2.211	27.669	0.267	15.668
16	9744	11585	13731	6952	2.084	27.595	0.566	15.632
18	9692	11555	13732	7033	1.933	27.545	0.981	15.589
20	9634	11521	13732	7120	1.76	27.527	1.529	15.537
22	9573	11484	13731	7213	1.57	27.551	2.215	15.477
24	9507	11444	13730	7311	1.376	27.632	3.031	15.41

Table IV-5. 3D Performance t_a = 9 Years Concluded

θ_a deg	J kg	m_a kg	m_D kg	P_D W	$-v_{H\rho}$ Gm/yr	$v_{H\phi}$ Gm/yr	v_{Hz} Gm/yr	$\Delta\phi$ rad
26	9438	11401	13728	7411	1.201	27.805	3.946	15.338
28	9366	11356	13724	·7513	1.098	28.173	4.893	15.265
30	9292	11308	13715	7610	1.185	29.042	5.715	15.2
32	9218	11257	13703	7700	1.477	30.442	6.131	15.153
34 ·	9145 ·	11209	13693	7791	1.677	31.5	6.28	15.121
36	9076	11164	13686	7883	1.769	32.183	6.352	15.095
38	9008	11121	13682	7974	1.801	32.646	6.395	15.072
40	8943	11079	.13679	8064	1.797	32.974	6.421	15.051
42	8880	11039	13677	8152	1.769	33.213	6.437	15.032
44	8818	11000	13675	8237	1.725	33.39	6.444	15.015
46	· 8758	10962	13674	8321	1.67	33.521	6.446	14.998
48	8700	10926	13673	8402	1.607	33.62	6.443	14.983
50	8643	10890	13672	8481	1.537	33.692	6.435	14.969
52	8587	10854	13672	8558	1.462	33.745	6.425	14.955
54	8533	10820	13672	8634	1.384	33.781	6.411	14.942
56	8480	10786	13671	8707	1.303	33.806	6.395	14.93
58	8429	10753	13671	8779	1.219	33.82	6.376	14.919
60	8377	10721	13671	8849	1.134	33.825	6.355	14.908
62	8327	10689	13671	8917	1.047	33.824	6.333	14.898
64	8278	10658	13671	8984	0.96	33.816	6.309	14.888
66	8230	10627	13672	9049	0.872	33.804	6.283	14.879
68	8182	10596	13672	9113	0.783	33.788	6.257	14.87
70	8136	10566	13672	9175	0.693	33.767	6.23	14.862
72	8090	10537	13672	9237	0.603	33.744	6.204	14.854
74	8045	10508	13673	9296	0.513	33.718	6.177	14.847
76	8001	10479	13673	9355	0.422	33.689	6.15	14.839
78	7957	10451	13673	9413	0.331	33.659	6.124	14.833
80	7914	10423	13674	9469	0.239	33.627	6.099	14.826
82	7872	10395	13674	9525	0.147	33.593	6.075	14.819
84	7830	10368	13675	9579	.054	33.559	6.051	14.813
86	7789	10340	13675	9633	.039	33.523	6.029	14.807
88	7748	10314	13675	9686	-0.132	33.487	6.008	14.801

	t (yr)	z (Tm)
1	0	0
2	0.28	-0.006
3	0.8	0.02
4	1.48	0.139
5	2.30	0.384
6	3.12	0.686
7	4.15	1.057
8	5.0	1.286

CLASS I TRAJECTORY

Figure IV-9. Three-Dimensional Trajectory for t_a = 5 Years, θ_a = 40 deg

	t (yr)	z (Tm)
1	0	0
2	0.28	0.001
3	0.62	0.008
4	1.01	0.002
5	1.34	-0.018
6	1.73	-0.019
7	2.31	0.04
8	3.28	0.236
9	4.82	0.683
10	7.00	1.286

CLASS II TRAJECTORY

Figure IV-10. Three-Dimensional Trajectory for t_a = 7 Years, θ_a = 40 deg

	t (yr)	z (Tm)
1	0	0
2	0.28	0.003
3	0.62	0.003
4	1.01	-0.008
5	1.36	-0.011
6	1.67	0.01
7	1.88	0.023
8	2.15	0.034
9	2.62	0.034
10	3.20	-0.001
11	3.78	-0.74
12	4.13	-0.12
13	4.37	-0.127
14	4.54	-0.094
15	4.69	-0.022
16	4.89	0.084
17	5.16	0.207
18	5.52	0.344
19	6.02	0.506
20	6.69	0.705
21	7.57	0.948
22	8.3	1.131
23	9.0	1.286

CLASS III TRAJECTORY

Figure IV-11. Three-Dimensional Trajectory for t_a = 9 Years, θ_a = 40 deg

Figure IV-12. Three-Dimensional Trajectory for t_a = 5 Years, θ_a = 80 deg

The figure contains the following data table:

	t (yr)	z (Tm)
1	0	0
2	0.14	-0.005
3	0.29	-0.008
4	0.48	-0.003
5	0.74	0.024
6	1.10	0.100
7	2.05	0.485
8	3.02	1.013
9	4.15	1.623
10	5.0	1.97

CLASS I TRAJECTORY

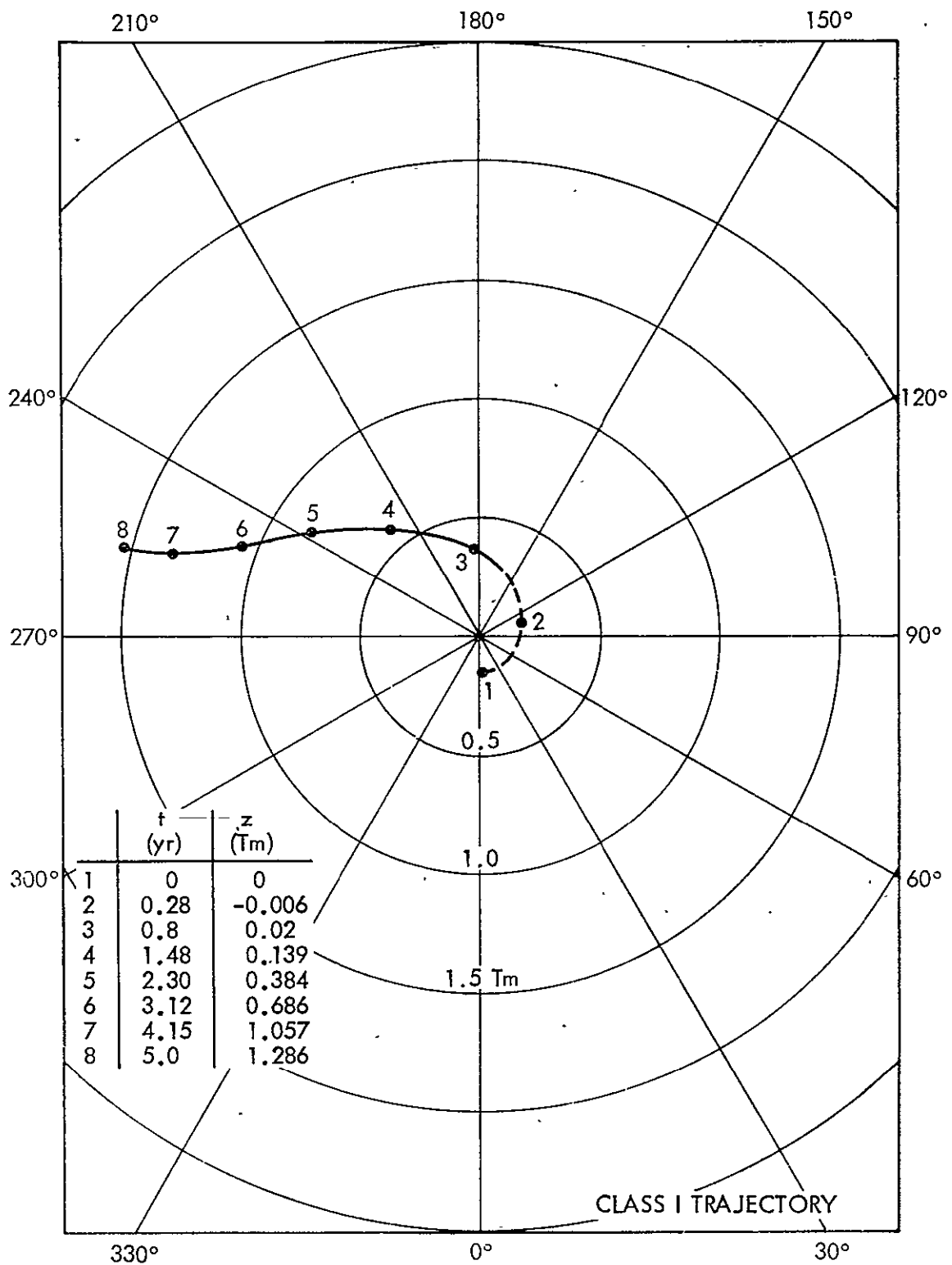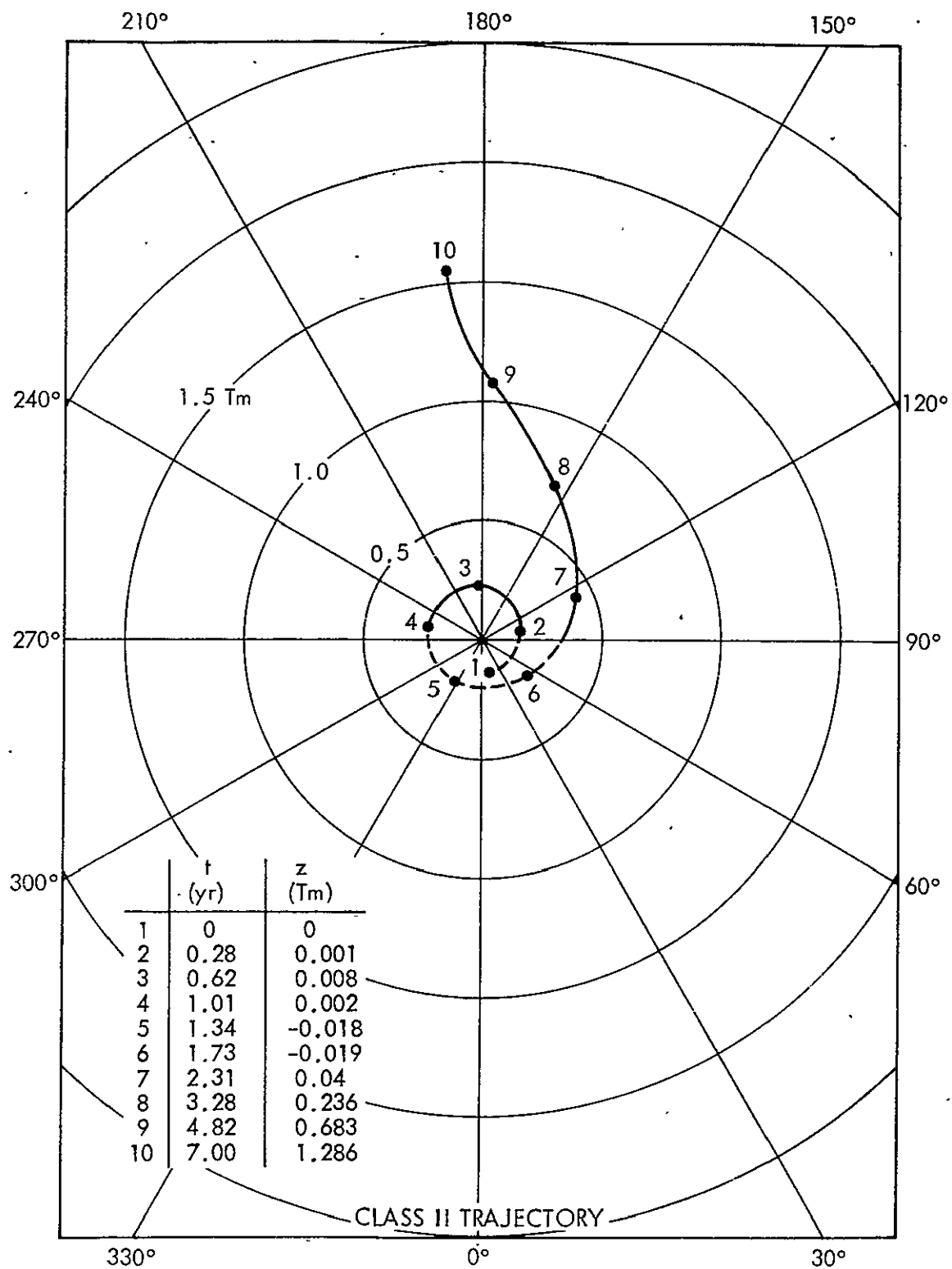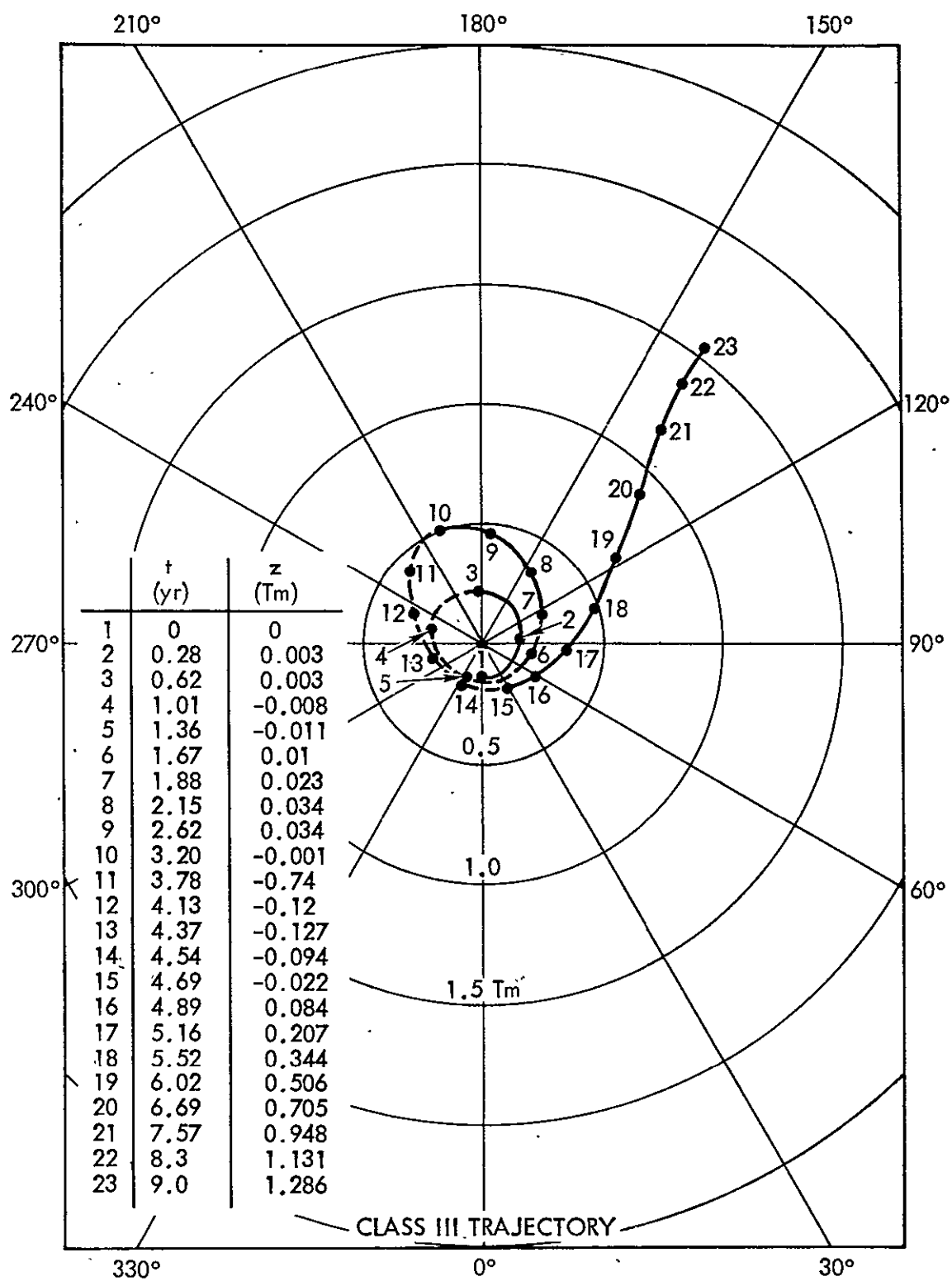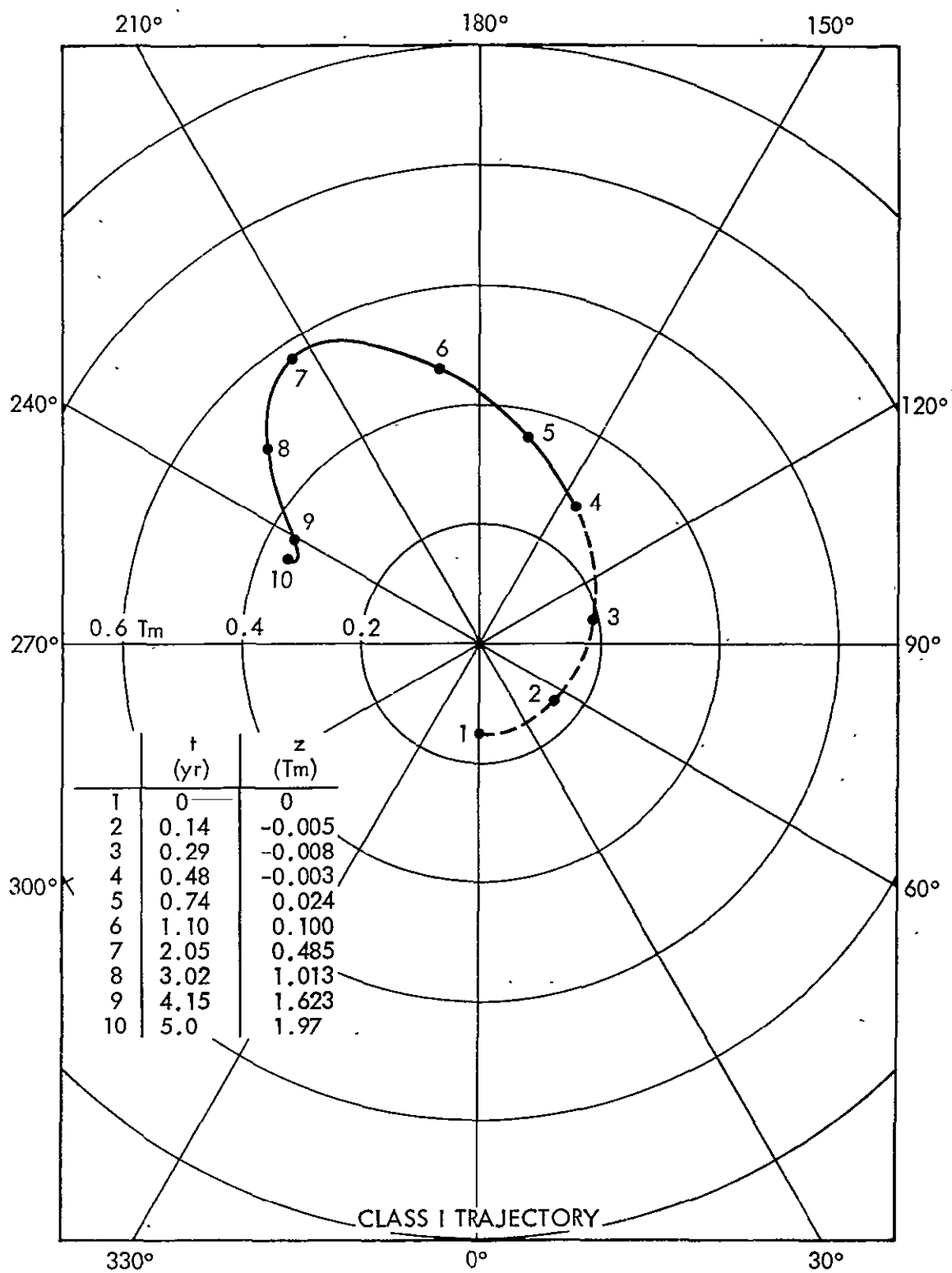|	t (yr)	z (Tm)
1	0	0
2	0.14	-0.002
3	0.28	-0.005
4	0.46	-0.006
5	0.69	-0.003
6	1.0	0.014
7	1.26	0.041
8	1.57	0.087
9	1.92	0.162
10	2.33	0.271
11	2.78	0.415
12	3.28	0.595
13	3.83	0.807
14	4.47	1.054
15	5.2	1.34
16	6.09	1.666
17	6.6	1.841
18	7.0	1.97

CLASS I TRAJECTORY
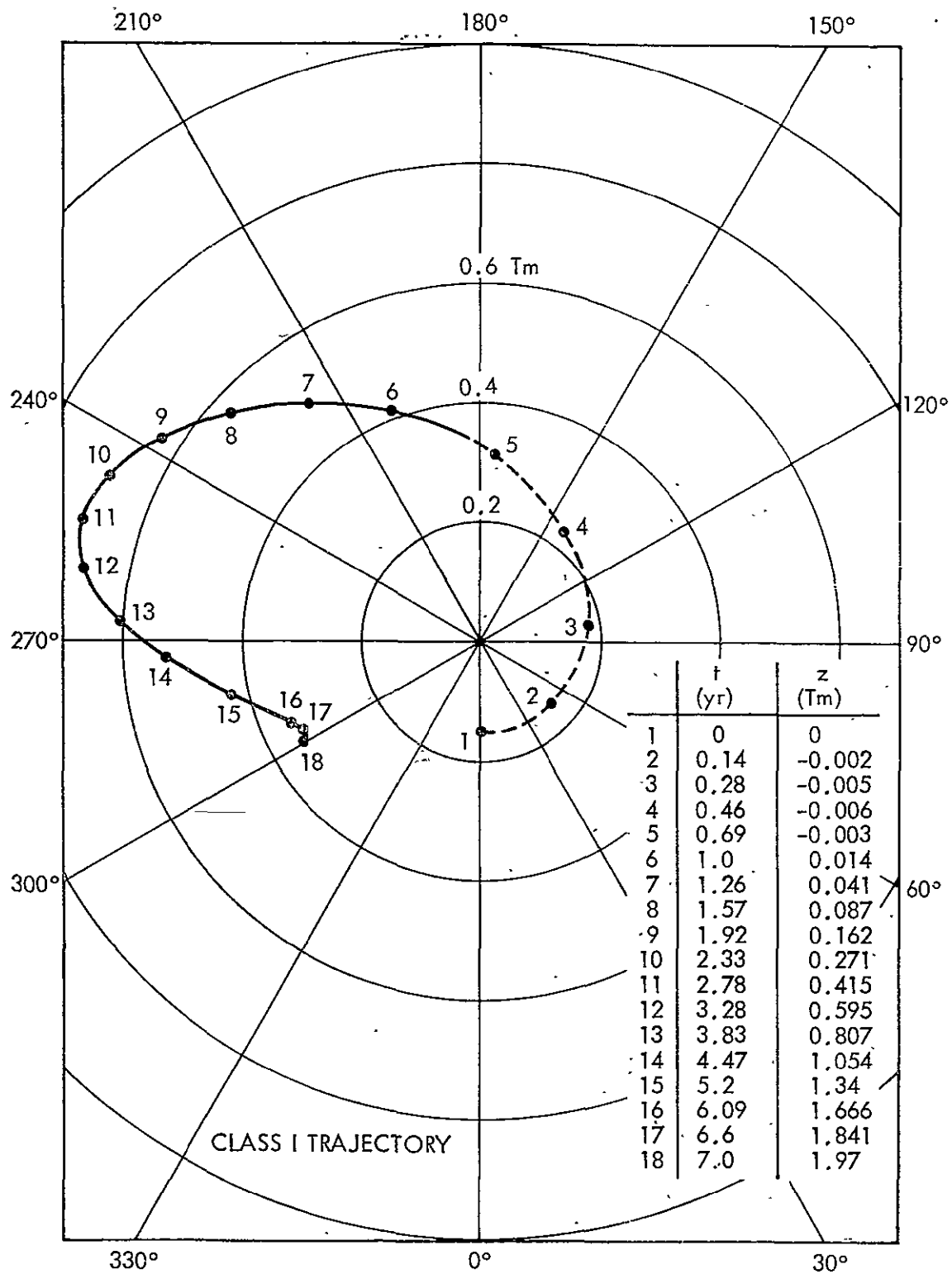
Figure IV-13. Three-Dimensional Trajectory for t_a = 7 Years, θ_a = 80 deg
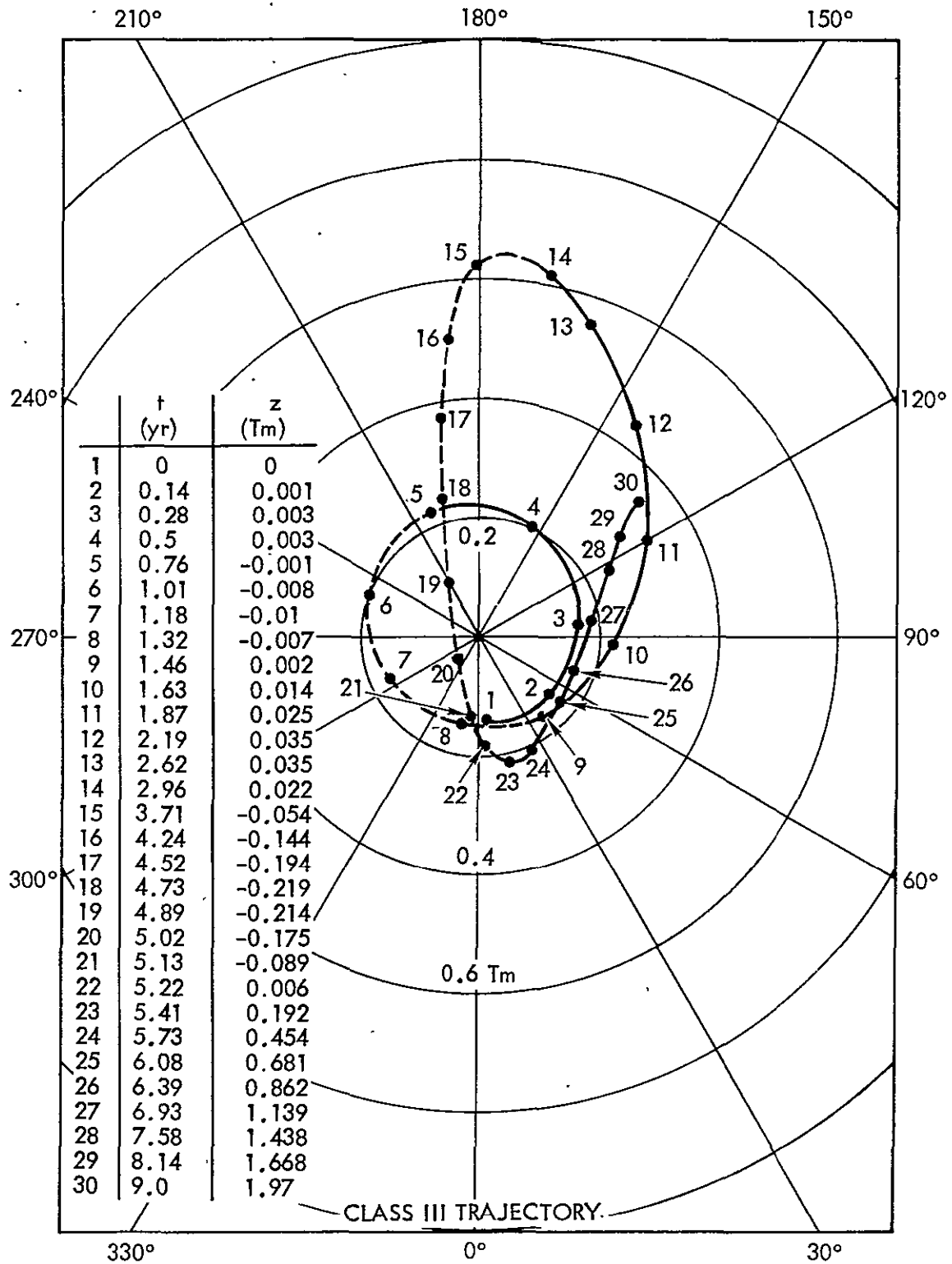
Figure IV-14. Three-Dimensional Trajectory for t_a = 9 Years, θ_a = 80 deg

The figure contains the following data table:

	t (yr)	z (Tm)
1	0	0
2	0.14	0.001
3	0.28	0.003
4	0.5	0.003
5	0.76	-0.001
6	1.01	-0.008
7	1.18	-0.01
8	1.32	-0.007
9	1.46	0.002
10	1.63	0.014
11	1.87	0.025
12	2.19	0.035
13	2.62	0.035
14	2.96	0.022
15	3.71	-0.054
16	4.24	-0.144
17	4.52	-0.194
18	4.73	-0.219
19	4.89	-0.214
20	5.02	-0.175
21	5.13	-0.089
22	5.22	0.006
23	5.41	0.192
24	5.73	0.454
25	6.08	0.681
26	6.39	0.862
27	6.93	1.139
28	7.58	1.438
29	8.14	1.668
30	9.0	1.97

CLASS III TRAJECTORY.

CHAPTER V

GRAVITATIONAL WAVE DETECTORS

1. Introduction

With spacecraft performance now established, we can turn to a discussion of the payload. Two basic types of gravitational wave detectors appear to have some promise for focal missions - what we will call compact and long wave detectors. The former have dimensions of a few meters at most, and are intended to cover frequencies in the kilohertz range. We will review them in the next section. The latter are intended for much lower frequencies, and are a great deal larger. To illustrate their problems and potential, we will concentrate on a design specifically intended to detect Crab Nebula pulsar radiation at 60.427 Hz; Sections 3-5 will cover this work.

The detection methods that we will discuss are all essentially strain gauges. The idea is that the spatial strain normal to the direction of propagation of a gravitational wave is, in principle, directly observable by a properly oriented, freely falling strain gauge. Numerous techniques for building detectors on this basic idea have been proposed. For an excellent summary see MTW,[66] Chapter 37. Several of these have been built, and others are under construction.

The main problems with these detectors all arise from the minute values of strain that we need to measure. For many detectors, the worst problem is thermal excitation (kT noise) of the vibrational modes of the gauge's mechanical foundation. Others are troubled by electrical noise in the transducer and following amplifiers. When these are mastered, quantum effects can play an important role. These arise when the strain variation is so weak that it cannot cause transitions between neighboring energy levels in the mechanical vibration modes. If this happens, the detector is transparent to the radiation.

Besides these fundamental problems, there are hosts of experimental difficulties. Perhaps the worst is vibration and acoustic isolation - the latter making high vacuum a necessity for all instruments known to the author. In addition, there is electromagnetic shielding, strain relief in every structural member under stress, and even the impacts of ambient nuclear particles. Finally, the measurement of such small strains, even in the absence of noise, is extremely challenging.

2.   Compact Detectors

The first serious experimental efforts were those of J. Weber and his colleagues at the University of Maryland. He used massive solid aluminum cylinders instrumented with axial strain gauges. Radiation propagating normal to the axis excites vibrations of the bar's axial modes. The idea here is that the thermal noise energy is of $O(kT)$ in each mode of the bar, independent of its mass M, where T is the bar's absolute temperature, and k is Boltzmann's constant. However, the amplitude due to this noise is proportional to $M^{-1/2}$, so heavy bars have lower thermal noise.

In his main work, Weber used bars 1.55 m long, 0.66 m in diameter, weighing about 1500 kg, and with a fundamental axial resonance of 1661 Hz. The arrangemen was most sensitive to radiation bursts with substantial energy content near this frequency. For details, see the bibliography in Ref. 66, and Weber's article in Ref. 11.

To lower the noise further, Weber actually employed two such antennas about 1000 km apart, and cross-correlated the outputs. In 1970-71,[92,93] he announced coincidences well in excess of chance, interpreting these as cosmic gravitational radiation bursts. Skepticism ensued, partly because such strong bursts were unexpected, but mainly because others following Weber's methods,

and with equal or better sensitivity, were unable to corroborate his results.[31], [29], [16], [54] Later, an error was found in his computer program, which, when fixed, reduced but did not eliminate his coincidences.[58] At this writing, the discrepancy and the controversy persist.

The development of Weber bars has been continued by several groups, of which we will mention three because some of their techniques may prove useful for focal missions. R. Drever, and his group at the University of Glasgow attempted to verify Weber's results. They used pairs of bars separated by piezoelectric transducers, and looked for both bursts[26] and continuous radiation.[48] Except for one plausible looking burst, their findings were essentially negative. More recently, this group has developed improved transducers based on laser interferometry.[32] By optical folding, the signal (and, unfortunately, the thermal noise) is magnified, so that the relative effect of electrical noise in the transducer and amplifier is reduced. To deal with thermal noise, they have attempted to increase the distance between their bars. In this they are limited by the stability of the bar's suspension, and by the expense of large vacuum chambers. We note here that high vacuum is free in space, and that the suspension problem takes on an entirely different character in free fall.

A direct attack on thermal noise is to cool the bars, transducers, and amplifiers to cryogenic levels. Three groups, working in concert, are following this idea, each building their own detectors. These are W. M. Fairbank, and his group here at Stanford,[40,41,69] W. O. Hamilton heading a group at Louisiana State University,[43] and a group including G. Pizzella, I. Modena, and S. V. Pallottino at the University of Rome.[71]

A range of different techniques is being tried by the three groups. The bar temperatures are planned to be in the range 0.003 - 0.05 K which will give thermal noise reductions by factors of $3 \times 10^3 - 10^5$. All will use superconducting magnetic suspensions. A variety of strain gauges and associated amplifiers is being tried. Stanford intends a superconducting quantum interference device (SQUID) which includes a Josephson junction amplifier.[69] The Romans are looking both at a low temperature ceramic piezoelectric tranducer with a FET amplifier, and at their own version of a SQUID magnetometer.[71] The LSU group is experimenting with a novel superconducting double cavity microwave accelerometer.[43] Amplifier noise is less of a problem here, because the output is the beat between the two cavities.

The development of this group of cryogenic techniques is particularly interesting for focal missions. The combination of great solar distance and long transit time means that temperatures of order 20 K should be reachable by direct radiation cooling, even in massive Weber bars. A further reduction to perhaps 0.01 K by refrigeration may then be feasible. This claim is not hard to substantiate. Consider a bar of mass M and specific heat C(T). For simplicity, suppose it is connected by a lossless heat pipe to a radiator of area A, and emissivity ε radiating to a 0 K universe. It then loses energy at a rate:

$$\dot{E} = \sigma \varepsilon A T^4 \qquad (1)$$

where $\sigma = 5.6697 \times 10^{-8}$ W/m^{2}K^4 = Stefan-Boltzmann constant. A loss dE causes a temperature drop dT given by:

$$dE = MC(T)dT \qquad (2)$$

Thus, the time to cool from T_2 to T_1 is:

$$t_1 - t_2 = \frac{M}{\sigma \varepsilon A} \int_{T_1}^{T_2} \frac{C(T)}{T^4} \, dT \qquad (3)$$

If we are given tabular values of $C(T)$, then, for the purpose of estimation, we can assume that $C(T)$ varies linearly between each pair of adjacent temperatures T_1 and T_2. In this approximation the integral becomes:

$$\int_{T_1}^{T_2} \frac{C(T)}{T^4} \, dT = \frac{1}{6 T_1 T_2} \left(\frac{1}{T_1} - \frac{1}{T_2} \right) \left[\left(1 + \frac{2 T_2}{T_1} \right) C(T_1) + \left(1 + \frac{2 T_1}{T_2} \right) C(T_2) \right] \qquad (4)$$

Supposing an aluminum bar, $C(T)$ has been taken from Ref. (45), and is shown in Table V-1. After converting to SI units, the integral has been evaluated between each pair of adjacent temperatures and is given in the third column. Finally, by taking illustrative values $M = 1000$ kg, $A = 1$ m^2, and $\varepsilon = 0.95$, we can compute the time for the temperature to fall between each pair of values. The results are shown in the last column.

Our example shows that a temperature of 23 K would be reached in only about a year. Below this temperature, our 0 K universe assumption begins to get shaky, and heat leaks from the warm parts of the spacecraft may become important. Nevertheless, even allowing for heat pipe losses, a radiation cooled heat sink at 15-20 K looks eminently feasible. We conclude that massive Weber bars, using very low temperature instrumentation, are possible payloads for electrically propelled focal missions.

Table V-1. Radiation Cooling Time

T	C(T)	$\frac{1}{\sigma}\int \frac{C(T)}{T^4} dT$	Δt
K	cal/g	m^2-sec/kg	years
23.16	.0039		
		7559	0.252
32.56	.0092		
		4111	0.137
40.16	.0165		
		11106	0.37
73.16	.076		
		4783	0.16
123.16	0.1367		
		1254	.042
173.16	0.1676		
		448	.015
223.16	0.1914		
		200	0.007
273.16	0.2079		
Totals		29461	0.983

A different way of improving Weber bar sensitivities is to reduce the damping. Each mode of the bar may be regarded as a damped second order resonato driven by thermal noise and external radiation. The damping does not affect either the average thermal energy kT in each mode or the energy deposited in that mode by the external input. However, the lower the damping, the lower is the likely rate of change of the thermal energy. Thus bursts of external radiation become (within limits) statistically more distinctive as the damping is lowered.

The cryogenic groups derive some benefit from this, as the acoustic damping in aluminum drops somewhat with temperature. However, Braginskii[15] and his group at Moscow State University have pointed out that the damping at any temperature is greatly reduced in very pure single crystals. They have been successful in fabricating sapphire crystals of 25 kg or more, and have already achieved sensitivities comparable with the more massive bars discussed above, although the damping in these crystals is still about two orders of magnitude higher than theoretical predictions. For transducers, the Moscow group appears to have worked mostly with capacitor types, and with the deformation of superconductive microwave cavity oscillators.

The importance of single crystal detectors for focal missions is obvious. With the greatly reduced bar masses, chemically powered spacecraft may be feasible. Using electrical propulsion, a fleet of detectors could be carried by a single spacecraft past the turn point, and then separated to compile a more comprehensive map of the focal area.

3. A Long Baseline Interferometer

The general idea of long baseline detectors is that if the strain field of a gravitational wave is uniform, a longer strain gauge will see greater displacements, without necessarily incurring increased noise. One way to do this is to analyze tracking data from interplanetary spacecraft taken by the NASA Deep Space Network. This appears to have been first proposed by A. J. Anderson[2],[3], and was given a relativistic analysis by Davies[25] and by Estabrook and Wahlquist.[38] An excellent discussion of the engineering difficulties in such measurements is given by J. D. Anderson[4], who, however, was primarily interested in relativistic effects on spacecraft orbits. The main problems are

.master oscillator stability, spacecraft and earth antenna dimensional stability, and propagation uncertainties in the interplanetary plasma.

The spacecraft tracking work is interesting here mainly because much of the analysis done there is generally applicable to all long baseline detectors. We will display some of it in order to draw a few conclusions for focal missions. To begin, several authors[66],[38] show that a weak gravitational wave in otherwise flat space propagating in the z direction, and polarized along x, can be described by the line element:

$$ds^2 = (1 + h)dx^2 + (1 - h)dy^2 + dz^2 - c^2dt^2 \tag{5}$$

where $h = h(t)$ is the metric disturbance.

Consider first a measurement of local strain in the x direction. For this $dy = dz = dt = 0$, so:

$$ds = \sqrt{1 + h}\, dx = \left[1 + \frac{h}{2} + O(h^2)\right] dx \tag{6}$$

The strain in the x direction is thus:

$$\varepsilon_x = \frac{ds - dx}{dx} \doteq h/2 \tag{7}$$

A similar analysis shows that

$$\varepsilon_y \doteq - h/2, \quad \varepsilon_z = \varepsilon_t = 0 \tag{8}$$

Thus there is no strain in the direction of propagation, and ideal clocks are unaffected by the wave.

If the goal is to detect this h, it seems intuitively obvious that the

baseline should be put along either x or y. Estabrook and Wahlquist[38] examined the more general case in which the spacecraft-tracking station line of sight lay in the x-z plane. For a sinusoidal wave:

$$h = h_0 s(\omega t) \tag{9}$$

they determined the resulting frequency shift $\Delta\nu/\nu$, since doppler measurements are superior to range for this purpose. However, for sinusoidal signals, the range and doppler amplitudes are proportional. As recently shown by the author[83], in an extension of their work, the guess that best performance comes from putting the baseline orthogonal to the wave direction is correct, provided L < 0.38964 λ. It is also shown that significant improvements are often possible at other angles, for larger L. Since we will not be concerned with baselines longer than λ/4, we will assume orthogonality from here on.

For this special case we can get the results we need directly from the line element. Suppose a packet of electromagnetic radiation is emitted at x = 0 at time t = t_e as measured by a local clock. Let it travel along the x axis to x = L, arriving there at coordinate time t = t_L. Then suppose it is reflected back, returning to x = 0 at t = t_r. On this path dy = dz = 0, and as the radiation travels a null geodesic, ds = 0 also. Thus, to first order in h, we can write on this path:

$$\frac{1}{c} dx = \pm(1 - \frac{h}{2})dt \tag{10}$$

Over the first leg this integrates to:

$$\frac{L}{c} = \int_{t_e}^{t_L} (1 - \frac{h}{2})dt = t_L - t_e + \frac{h_0}{2\omega}\left[c(\omega t_L) - c(\omega t_e)\right] \tag{11}$$

Similarly, on the return leg:

$$\frac{L}{c} = \int_{t_L}^{t_r} (1 - \frac{h}{2})dt = t_r - t_L + \frac{h_0}{2\omega}\left[c(\omega t_r) - c(\omega t_L)\right] \tag{12}$$

Adding these eliminates the unobservable t_L:

$$\frac{2L}{c} = t_r - t_e + \frac{h_0}{\omega} s\left[\frac{\omega}{2}(t_r - t_e)\right]s\left[\frac{\omega}{2}(t_e + t_r)\right] \tag{13}$$

Letting $t_m = \frac{1}{2}(t_e + t_r)$ be the average time of the measurement, then to first order in h_0 this is:

$$\frac{2L}{c} = t_r - t_e + \frac{h_0}{\omega} s(\frac{\omega L}{c})s(\omega t_m) \tag{14}$$

Interpreting the difference $t_r - t_e$ as a measure of the round trip distance, the variation in L due to h is:

$$\Delta x = \frac{c}{2}(t_r - t_e) - L = \frac{h_0 c}{2\omega} s(\frac{\omega L}{c})s(\omega t_m) \tag{15}$$

Finally, substituting the wavelength of the gravitational radiation:

$$\lambda = \frac{2\pi c}{\omega} \tag{16}$$

we get:

$$\Delta x = \frac{h_0 \lambda}{4\pi} s(\frac{2\pi L}{\lambda})s(\omega t_m) \tag{17}$$

Note that for $L \ll \lambda$, $t_m = t$, and $\Delta x = hL/2$, consistent with our earlier result that the strain is h/2. This is what we are attempting to measure with Weber bars.

The most important result from this is that the amplitude of Δx is maximized at $L = \lambda/4$. Actually, except for very small λ, measurement difficulties

increase with L, so engineering considerations will generally lead to L < λ/4. On the other hand, we will later show that L = λ/4 exactly has an interesting advantage in reducing laser noise. For the Crab pulsar λ = 4961 km, and the optimum baseline would be 1240 km. If we observe this baseline optically with, say, green light of wavelength 5×10^{-7} m, and are capable of directly measuring 1% of this, then the minimum directly observable h_0 is

$$h_0 = \frac{4\pi(5 \times 10^{-7})(.01)}{(4.961 \times 10^{6})(1)(2)} = 6.3 \times 10^{-15}$$

Actually, since the optical signal from the Crab is also available, we can employ cross-correlation over a long time to improve detection. If, say, 10 days are devoted to this, then $N = (8.64 \times 10^{5})(60.427) = 5.22 \times 10^{7}$ cycles are available, and we get a detection gain of $\sqrt{N} = 7226$. This lowers the minimum detectable h_0 to 8.8×10^{-19}. Not good enough, according to most estimates of pulsar emissions.

Further improvements are possible, for which we must first examine the optical link. Suppose we have a laser source with wavelength $\lambda_L = 5 \times 10^{-7}$ m, as above. Let its output be directed by a parabolic mirror of diameter D = 0.4 m. Then, the half beamwidth is $\Theta_W = 1.22\ \lambda_L/D = 1.53 \times 10^{-6}$ rad., and the radius to first null at the other end is $\Theta_W L = 1.89$ m. This is the reason for not using a larger directive mirror - the collector will already be a significant fraction of the pattern size, and directive pointing of the order of 2×10^{-7} rad. is already required.

The tight beam makes it possible to use the simplest kind of collector at the far end - a corner cube. This avoids both pointing problems and phase stability problems in laser amplifiers. Since we must be careful of interference

effects, we must use a single large cube, unlike the arrays left on the moon by the Apollo astronauts. To establish the mass and optical properties, the following geometrical relations are easily proved. If the corner is cut from a cube of edge a, then

$$\text{Volume} = V = a^3/6 \tag{18}$$

$$\text{Frontal area} = A = \sqrt{3/4}\, a^2 \tag{19}$$

$$\text{"Effective" diameter} = D_c = \sqrt{\frac{4A}{\pi}} = \sqrt{\frac{2\sqrt{3}}{\pi}}\, a = 1.05a \tag{20}$$

The effective diameter is defined, as usual, as the diameter of a circle of area A.

Since pointing is not a problem, the limiting size of the reflector is set by its mass. Suppose we allow, say, 50 kg of high transparency lead glass (3000 kg/m^3). This gives:

$$a = (6V)^{1/3} = \left[\frac{(6)(50)}{3000}\right]^{1/3} = 0.464 \text{ m}$$

from which A = 0.187 m^2 and D_c = 0.487 m. The pattern back at the source will not be circular, but it will have a rough radius 1.22 λ_L/D_c = 1.55 m. This is adequate for any collecting aperture we are likely to use. An actual corner cube of the same mass would probably be a little deeper and have its frontal corners trimmed. Finding the trim shape that would give the strongest return beam for a given mass of glass is an interesting problem in optimization theory, but it's doubtful that much improvement could be found over the present figures.

We can now calculate the fraction F of the initial laser output that is collected at the detector. First, the gain of a tapered parabolic reflector is:[73]

$$G = \frac{7\pi}{4}\left(\frac{D}{\lambda_L}\right)^2 \tag{21}$$

so the fraction intercepted by the corner cube is

$$\frac{GA}{4\pi L^2} = \frac{7A}{16}\left(\frac{D}{\lambda_L L}\right)^2 = \frac{(7)(0.187)}{16}\left[\frac{0.4}{(5 \times 10^{-7})(1.24 \times 10^6)}\right]^2 = .0340$$

Assuming a final collecting aperture of 0.2 m^2 (0.505 m diameter paraboloid, say), the equivalent return fraction is:

$$\frac{(7)(0.2)}{16}\left[\frac{(0.487)}{(5 \times 10^{-7})(1.24 \times 10^6)}\right]^2 = .0541$$

Combining these, and allowing for a 3% light loss on each of ten surfaces, we get an overall F = 1.36 x 10^{-3}.

The next question is what to do with the returned beam. If we attempt to detect the change Δx in L by comparing it with the phase of the current laser output, then we place extremely tight demands on the phase stability of the laser over the round trip travel time (.0083 sec in this case). A way around this is to use an interferometric setup, such as in Fig. V-1. Here the laser is split into two beams which are sent to separate corner reflectors along the x and y directions, each at the distance L. When the beams are recombined, there is no differential time delay, and laser phase stability is not so important. We will now set out to examine the performance of such a system.
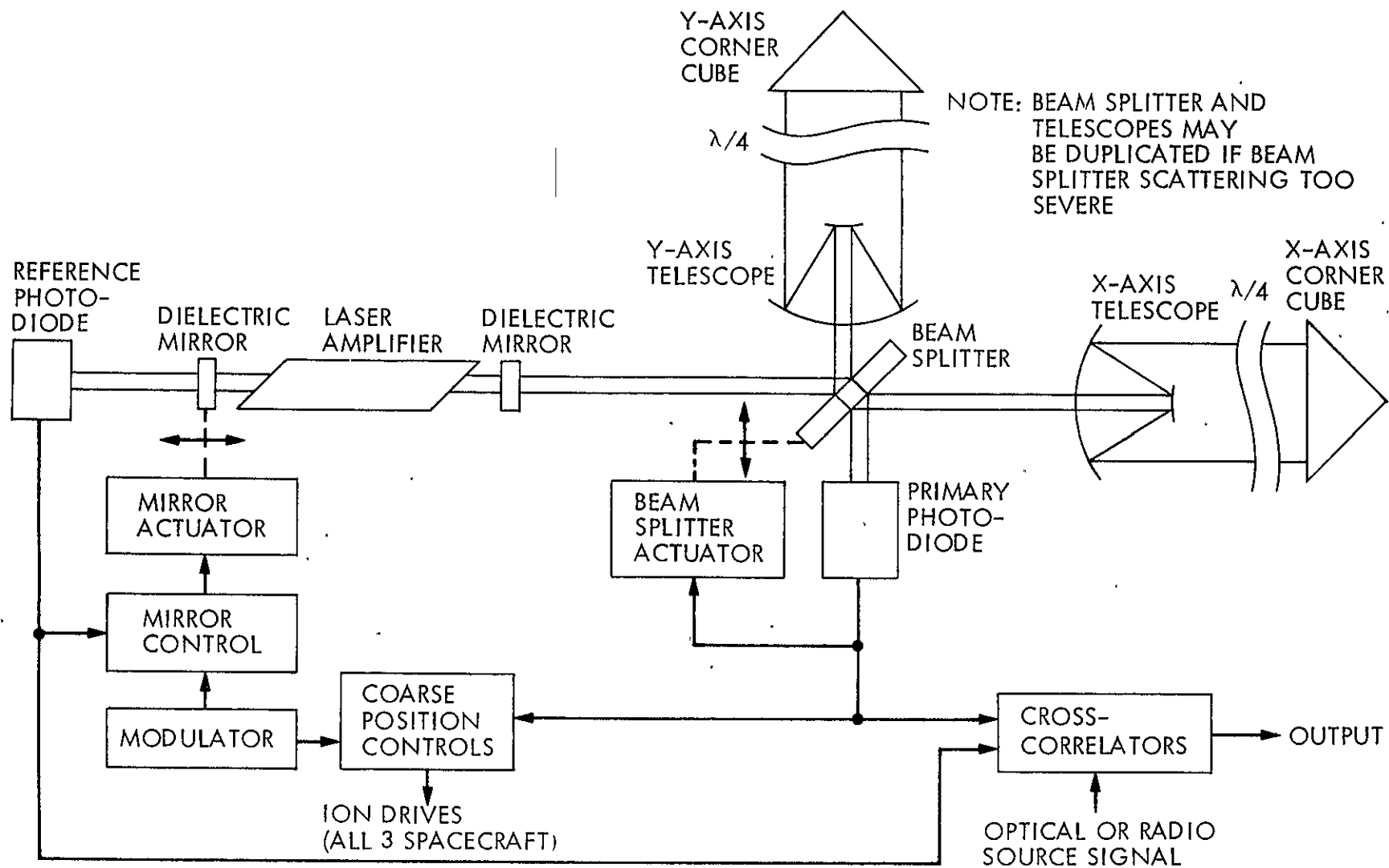
Figure V-1. A Space Tuned Long Wave Detector

4. Spatial Filtering

To begin our interferometer analysis, we must characterize the output of the laser. Assuming a linearly polarized wave, the nominal electric field will have the form $E_0 s(\omega_L t)$. However, practical lasers (even single mode types) are contaminated by spontaneous emission noise, and frequency instability due to acoustic vibration of the resonant cavity. Spontaneously emitted photons lead to a field component at a frequency close to ω_L, but with random phase. The effect is an additive noise spectrum, centered at ω_L, of very narrow width (see Ref. 98, Sec. 10.6).

As for acoustic noise, imagine a weak axial vibration of the laser struc- ture at frequency ω_m. By changing the mirror spacing, it will frequency modu- late the laser. If the maximum frequency excursion is $\Delta\omega$, then the output field will have the form:

$$E = E_0 s\left[\omega_L t + gs(\omega_m t)\right] \tag{22}$$

where $g \equiv \Delta\omega/\omega_m$. (See Ref. 73, P. 532.) Expanding this in Bessel functions gives:

$$E/E_0 = J_0(g)s(\omega_L t) + \sum_{i=1}^{\infty} J_i(g)\left[s(\omega_L + i\omega_m)t + (-1)^i s(\omega_L - i\omega_m)t\right] \tag{23}$$

If this wave is incident on a detector, the output·current will depend on E^2. After averaging over the optical frequencies, this becomes:

$$\overline{(E/E_0)^2} = \frac{1}{2} J_0^2(g) + 2J_0(g) \sum_{i=1}^{\infty} J_i(g)c(2i\omega_m t)$$

$$+ \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} J_i(g)J_j(g)\left[c(i-j)\omega_m t + (-1)^i c(i+j)\omega_m t\right] \quad (i+j = \text{even}) \quad (2)$$

We conclude that the detector will see only even harmonics of the acoustic frequency ω_m.

The practical effect of this is curious. If we choose to eliminate the signal frequency ω from the laser output power by controlling the mirror spacing, then it is necessary to introduce an even subharmonic of ω of appropriate amplitude and phase into one mirror. The design of a controller to do this looks like a very interesting problem in optimal control; but it is doubtful that such a technique by itself could deliver a sufficiently pure signal to allow detection of pulsar radiation.

It doesn't take much acoustic strain to produce these problems. The complicated spectrum of FM harmonics just seen occurs when $g \gtrsim 1$. Thus to reduce the problem to only a few weak harmonics of ω_m, it is desirable to make $g \ll 1$. Taking $\omega_m = \omega/2$ as the worst offending acoustic frequency, this means we require $\Delta\omega \ll \omega/2$. Now the laser frequency ω_L depends inversely on the mirror spacing ℓ. Thus

$$\Delta\omega = -\omega_L \frac{\Delta\ell}{\ell} \tag{25}$$

and the allowable strain is

$$\varepsilon = \frac{\Delta\ell}{\ell} \ll \frac{\omega}{2\omega_L} = \frac{f\lambda_\ell}{2c} = \frac{(60.427)(5 \times 10^{-7})}{(2)(3 \times 10^8)} = 5 \times 10^{-14}$$

If we imagine that the mirrors are masses in relative oscillation with amplitude $\Delta \ell = \epsilon \ell$, then the displacement is

$$x = \epsilon \ell s(\tfrac{1}{2}\omega t) \qquad (26)$$

and the total energy of the vibration is

$$T.E. = \frac{1}{2} m(\dot{x}_{max})^2 = \frac{1}{8} m(\epsilon \omega \ell)^2 = \frac{1}{2} m(\pi \epsilon f \ell)^2 \qquad (27)$$

If 1 kg of structure is allotted to each mirror, and 1 m separation is used, then at the above strain limit:

$$T.E. = \frac{1}{2} (2)\left[\pi(5 \times 10^{-14})(60.427)(1)\right]^2 = 9 \times 10^{-23} j$$

Thus acoustic energy of the order of normal thermal excitation ($\sim kT/mode$) is already very troublesome. We conclude that elaborate measures for acoustic isolation and damping may be required.

Disturbances other than acoustic noise (e.g., spontaneous emission, spurious laser modes) take the form of additions to the electric field. Consider a component of this noise at frequency ω_m away from the main output. The field is then:

$$E = E_0 s(\omega_L t) + \epsilon s\left[(\omega_L + \dot{\omega}_m)t + \xi\right] \qquad (28)$$

The detector current due to this field would depend on:

$$\overline{E^2} = \frac{1}{2} E_0^2 + \frac{1}{2} \epsilon^2 + \epsilon E_0 c(\omega_m t + \xi) \qquad (29)$$

Thus the difference frequency ω_m now appears directly in the output. Reduction of this disturbance along with the acoustic noise by the above con-

troller looks just as feasible.

Further reduction of these disturbances is possible by a sort of spatial filtering. Suppose the current in the final detector contains a noise component at a frequency very close to the expected signal frequency ω. This will have the form $Ks(\omega + \varepsilon)t$. If a portion of the laser output is monitored directly, the corresponding current component will have the form $K's(\omega + \varepsilon)(t + \frac{2L}{c})$, where L is again the optical baseline. Adding these signals, and assuming some sort of control loop capable of forcing $K' = K$, the combined current from this noise component is:

$$I/K = s(\omega + \varepsilon)t + s\left[(\omega + \varepsilon)(t + \frac{2L}{c})\right]$$

$$= 2s\left[(\omega + \varepsilon)(t + \frac{L}{c})\right]c\left[\frac{L}{c}(\omega + \varepsilon)\right] \tag{30}$$

Now suppose L differs from a quarter wave only slightly:

$$L = \frac{\lambda}{4} + \delta = \frac{\pi c}{2\omega} + \delta \tag{31}$$

Then to first-order in ε and δ we get:

$$I = -K(\frac{\pi\varepsilon}{\omega} + \frac{2\omega\delta}{c})s\left[(\omega + \varepsilon)(t + \frac{L}{c})\right] \tag{32}$$

Thus, the disturbance can be cancelled out down to a level dependent on ε and δ. Consider the frequency first. Ultimately, the correlation detector will reject everything outside a band $\Delta f = 1/T$. For a measurement time T = 10 days, we have $\Delta f = 1.157 \times 10^{-6}$ Hz. Thus the largest angular frequency deviation is

$$\varepsilon_{max} = 2\pi(\frac{1}{2}\Delta f) = \pi\Delta f \qquad (33)$$

and, for sufficiently small δ, the noise within Δf is reduced to a fraction which is at worst:

$$\frac{\pi\varepsilon_{max}}{\omega} = \frac{\pi\Delta f}{2f} = \frac{\pi(1.157 \times 10^{-6})}{(2)(60.427)} = 3 \times 10^{-8}$$

In order to achieve so deep a cancellation, the optical baseline error δ must be rigidly controlled. At the same error level:

$$|\delta| = \frac{\pi\varepsilon_{max}c}{2\omega^2} = \frac{\Delta fc}{8f^2} = \frac{(1.157 \times 10^{-6})(3 \times 10^8)}{(8)(60.427)^2} = 0.012 \text{ m}$$

Thus, for best performance, the baseline should be a quarter wave to within 1-2 mm. A startling requirement.

Overall, we have seen that laser instability, particularly FM due to acoustic vibration, is a severe problem. While acoustic isolation, damping, mirror control, and spatial filtering all help to reduce the problem, it remains to be seen whether any combination of these or other measures will be sufficiently effective to permit detection of gravitational radiation from pulsars.

5. Detection Sensitivity

We turn now to detector problems. First let's look at the phase shift due to the gravitational wave. Over the round trip x axis path, the optical phase shift is, from (17):

$$\Delta\phi = \frac{4\pi\Delta x}{\lambda_L} = \frac{h_o\lambda}{\lambda_L}s(\frac{2\pi L}{\lambda})s(\omega t) = \frac{h_o\lambda}{\lambda_L}s(\omega t) \qquad (34)$$

in which we have assumed that $L = \lambda/4$. The same reasoning applied to the y axis leg gives a corresponding phase shift $-\Delta\phi$. However, different return strengths for the two optical signals are to be expected. Supposing the individual electric field strength amplitudes to be E_x and E_y, then the combined field at the detector is:

$$E = E_x s(\omega_L t + \Delta\phi) + E_y s(\omega_L t - \Delta\phi + \alpha) \tag{35}$$

where α is the phase shift due to the slightly different lengths of the two baselines.

Since detectors respond to power, we square and average over the optical frequencies:

$$\overline{E^2} = \frac{1}{2}E_x^2 + \frac{1}{2}E_y^2 + E_x E_y (c\alpha + 2\Delta\phi s\alpha) \tag{36}$$

Now the return power in the x-beam is $P_L F_x/4$, where P_L is the laser output power, F_x is the return fraction computed above, and the 4 is to account for the beam splitter and combiner (see Figure V-1). After allowing for asymmetries in the latter, and in the optical components, we can write the returned powers as:

$$P_x = \frac{1}{4}P_L F(1 + \mu); \quad P_y = \frac{1}{4}P_L F(1 - \mu) \tag{37}$$

where an exact match would make $\mu = 0$. Now since power is proportional to E^2, the optically averaged power delivered to either detector is:

$$P_D = \frac{1}{4}P_L F\left[1 + \sqrt{1-\mu^2}\,(c\alpha + 2\Delta\phi s\alpha)\right] \tag{38}$$

For any optical detector we are likely to consider, the output current can be taken as:

$$I = I_0 + \frac{e\eta\lambda_L P_D}{h_p c} \tag{39}$$

where e is the electronic charge, h_p is Planck's constant, and η is the quantum efficiency of the detector. Combining (38) and (39), we can write the detector

current as:

$$I = I_o + I_c[1 + \gamma(c\alpha + 2\Delta\phi s\alpha)] \tag{40}$$

where

$$\gamma \equiv \sqrt{1 - \mu^2} \tag{41}$$

and

$$I_c \equiv \frac{en\lambda_L P_L F}{4h_p c} \tag{42}$$

The latter is the nominal detector current, were we to set $\alpha = \pi/2$. Suppose we take the laser output power to be $P_L = 10$ watts. Then assuming a silicon photodiode with a quantum efficiency $\eta = 0.3$, and the previous numbers:

$$I_c = \frac{(1.6021 \times 10^{-19})(0.3)(5 \times 10^{-7})(10)(1.36 \times 10^{-3})}{(4)(6.6256 \times 10^{-34})(2.9979 \times 10^{8})} = 4.11 \times 10^{-4} \text{ A}$$

This is much larger than typical photodiode dark current values $I_o \sim 10^{-9}$ A.

We will now consider various noise sources in order to find good values for α and γ. Starting with detector shot noise, the rms noise current within the frequency band Δf for any kind of detector is (Ref. 98, p. 281):

$$\overline{I_{SN}} = \{2e\Delta f [I_o + I_c (1 + \gamma c\alpha)]\}^{\frac{1}{2}} \tag{43}$$

Thus the ratio of signal to rms noise is:

$$\frac{I_{SIG}}{I_{SN}} = \gamma\Delta\phi \sqrt{\frac{2I_c s^2\alpha}{e\Delta f(1 + \varepsilon + \gamma c\alpha)}} \tag{44}$$

where

$$\varepsilon \equiv I_o/I_c \tag{45}$$

For $\varepsilon = 0$ and $\gamma = 1$, this reduces to:

$$\frac{I_{SIG}}{I_{SN}} = \Delta\phi \sqrt{\frac{2I_c}{e\Delta f} (1 - c\alpha)} \tag{46}$$

whose maximum occurs at $\alpha = \pi$. We are led to the curious conclusion that to maximize the signal to noise ratio, we have to throw away the signal! However, the signal to noise ratio is not well behaved near this point. This is best seen in Fig. V-2, which is a plot of the function

$$Q(\alpha) = \gamma s\alpha(1 + \epsilon + \gamma c\alpha)^{-\frac{1}{2}} \tag{47}$$

for several values of ϵ and μ (i.e. γ).

In practice, it should not be hard to achieve $\epsilon < 10^{-3}$ by photodiode selection. However, matching return gains to better than $\mu = 0.3$ may be difficult without some sort of amplitude modulator and control loop. Lacking these, it seems prudent to choose $\alpha \doteq 135$ deg so that for $\mu \leq 0.3$, we have $Q \geq 1.18$. Addition of the control can raise this only 20% at most; so it's probably not worth the trouble.

We can now calculate the sensitivity. The rms signal phase variation is, from (34):

$$\Delta\phi_{rms} = \frac{h_0\lambda}{\sqrt{2}\lambda_L} \tag{48}$$

from which the ratio of rms signal to rms shot noise is, from (44):

$$\frac{I_{SIG\ rms}}{I_{SN}} = \frac{h_0\lambda Q}{\lambda_L}\sqrt{\frac{I_c}{e\Delta f}} \tag{49}$$

The usual criterion for detectability is that this ratio is at least unity. Thus the minimum detectable h_0 is:

$$h_0 = \frac{\lambda_L}{\lambda Q}\sqrt{\frac{e\Delta f}{I_c}} \tag{50}$$

and using the previous numbers, we could detect the Crab pulsar provided

$$h_0 > \frac{5 \times 10^{-7}}{(4.961 \times 10^6)(1.18)}\sqrt{\frac{(1.6021 \times 10^{-19})}{(8.64 \times 10^5)(4.11 \times 10^{-4})}} = 1.81 \times 10^{-24}$$
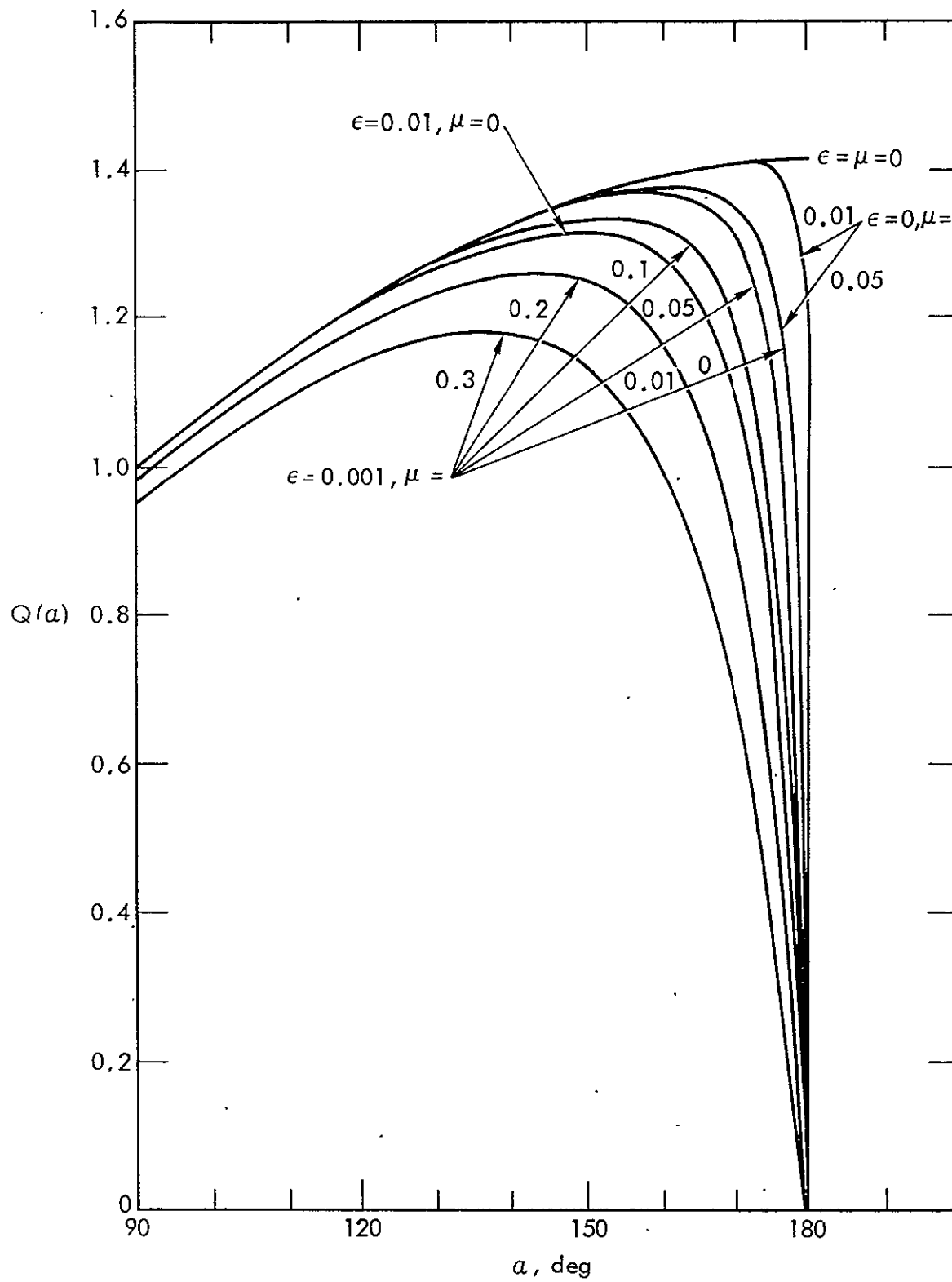
Figure V-2. Shot Noise Function for Various Values of μ and ϵ

We can now look at thermal noise. To couple the photodiode to an amplifier efficiently, the amplifier input resistance should be close to the diode's operating resistance. The latter is, from (40), (45), and (47):

$$R = \frac{V}{I_0 + I_c(1 + \gamma c\alpha)} = \frac{V}{I_c}\left(\frac{\gamma s\alpha}{Q}\right)^2 \tag{51}$$

where V is the reverse bias applied to the diode, typically about 20 volts. With the earlier numbers this gives us an amplifier input resistance of R = 1.3 x 10^4 ohms, an easy value to implement.

The rms noise current generated in this resistor, in the frequency band Δf, is

$$I_{TN} = \sqrt{\frac{4kT\Delta f}{R}} = \frac{2Q}{\gamma s\alpha}\sqrt{\frac{kTI_c\Delta f}{V}} \tag{52}$$

where T is the absolute temperature of the resistor, and k = 1.3805 x 10^{-23} J/K is Boltzmann's constant. Applying the same argument as for shot noise, the ratio of rms signal to rms thermal noise is:

$$\frac{I_{SIG\ rms}}{I_{TN}} = \frac{h_0\lambda\gamma^2 s^2\alpha}{Q\lambda_L}\sqrt{\frac{VI_c}{2kT\Delta f}} \tag{53}$$

Thus, if thermal noise were the only problem, the minimum detectable h_0 would be:

$$h_0 = \frac{Q\lambda_L}{(1 - \mu^2)s^2\alpha}\sqrt{\frac{2kT\Delta f}{VI_c}} \tag{54}$$

and assuming that the resistor is cooled to, say, 50 K, this works out to:

$$h_0 = \frac{(1.18)(5 \times 10^{-7})}{(4.961 \times 10^6)(1 - 0.3^2)(.5)}\sqrt{\frac{(2)(1.3805 \times 10^{-23})(50)}{(8.64 \times 10^5)(20)(4.11 \times 10^{-4})}} = 1.15 \times 10^{-25}$$

We conclude that thermal noise is an order of magnitude lower than shot noise, even though α has been optimized for the latter.

Finally, we must consider laser noise. Suppose a fraction δ of the original power P_L is noise in sidebands of width Δf around the main laser frequency. Then the rms noise current due to this is

$$I_{LN} = I_c \delta (1 + \gamma c \alpha) \tag{55}$$

This time, the signal to rms noise ratio is

$$\frac{I_{SIG}}{I_{LN}} = \frac{2\gamma \Delta \phi s \alpha}{\delta (1 + \gamma c \alpha)} \tag{56}$$

In this case, the sensitivity depends on the function

$$W(\alpha) = \gamma s \alpha (1 + \gamma c \alpha)^{-1} \tag{57}$$

Once again, if $\gamma = 1$, we find that $\alpha = \pi$ is optimum, when $W = \infty$; i. e., we can get an infinite signal to noise ratio by throwing away the signal! Of course, in the neighborhood of $\alpha = \pi$, $W(\alpha)$ shows even worse behavior than $Q(\alpha)$, which plays the same role in the shot noise analysis. This is best seen in Fig. V-3, which plots $W(\alpha)$ for various values of γ. If we again choose $\alpha = 135$ deg., consistent with the shot noise discussion, we find that for $\mu < 0.3$, $W > 2.07$.

Once again we form the rms signal to rms noise ratio:

$$\frac{I_{SIGrms}}{I_{LN}} = \frac{\sqrt{2}\ h_0 \lambda W}{\lambda_L \delta} \tag{58}$$

Thus, in order to detect a given h_0, we require

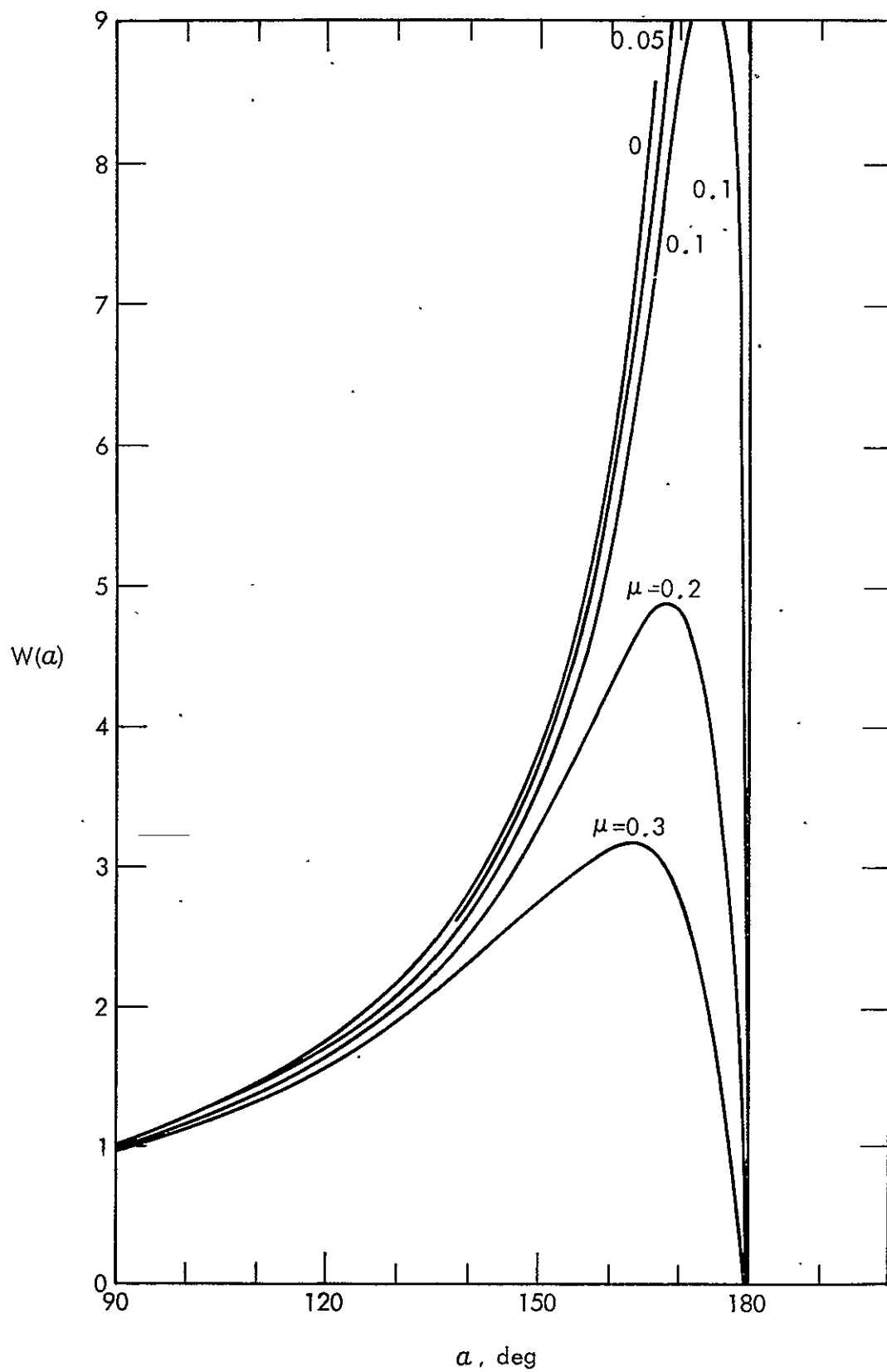$$\delta < \frac{\sqrt{2}\ h_0 \lambda W}{\lambda_L} \tag{59}$$

Figure V-3. Laser Noise Function for Various Values of μ

and if we wish the laser noise problem to be no worse than the shot noise, we must have

$$\delta < \frac{\sqrt{2} \ (1.81 \times 10^{-24})(4.961 \times 10^{6})(2.07)}{5 \times 10^{-7}} = 5.27 \times 10^{-11}$$

This means that through acoustic isolation, mirror control, and spatial filtering, no more than this fraction of P_L in the critical sidebands may survive.

In the author's opinion, a practical system with $\delta < 10^{-12}$, say, is feasible; but we have not proved this. In case of difficulty, however, we have the option of adding the return beam matching control, as discussed above. This would greatly reduce μ, so that α could be increased. An increase of $W(\alpha)$ to 15 or 20 may then be possible.

In conclusion, we have presented a long baseline laser interferometer, capable of detecting gravitational radiation at the Crab pulsar double frequency of 60.427 Hz, provided that the metric perturbation in the solar system $h_0 > 1.81 \times 10^{-24}$. This limit is set primarily by shot noise in the photodetector.

Several difficult control problems have been identified, but not addressed. One of these is the baseline control, which must establish each baseline length L to within a few millimeters of a quarter wave (1240 km). Another is the differential phase control, which must hold the baseline difference to a fairly precise optical phase α. In addition there are the acquisition and attitude controls for all three spacecraft, although the corner reflector controls do not look difficult. Laser and gain matching controls may also be needed to keep down laser noise.

Maintaining the required three spacecraft configuration in low earth orbit does not look very practical; so this must be considered to be a deep space system. In this connection we may note that configuration maintenance problems are least when the gravity force vector is the same for all three spacecraft. This condition is met almost exactly for focal missions. Finally, we observe that

while no mass estimate has been made for this system, the three spacecraft might just fit within the capability of the chemically powered injection vehicles of Chapter III.

LOOKING AHEAD

1.   Introduction

This thesis has been mainly concerned with a description of the solar foci,
and how to reach them.   This is, at best, a beginning; as many important aspects
of focal missions are untouched, or only lightly covered.   It is the author's
hope that future physicists and engineers will be moved to continue this work.
Accordingly, this final chapter provides a brief sketch of some of the remaining
areas bearing on mission design.

Section 2 provides a discussion of presently plausible sources of gravita-
tional radiation.   In Section 3, the way is opened for a more comprehensive lens
analysis than that in Chapter II, although the more difficult problems remain
unsolved.   Section 4 looks at the main navigational question - the feasibility
of following the axis precisely, beyond the turn point, despite our inability
to sense it directly.   Finally, in Section 5, the several diverse threads of this
thesis are combined to construct a set of (rather speculative) focal mission
profiles.   Many unsolved problems are identified for future study.

2.   Radiation Sources

Many suggestions have been made of plausible cosmic sources of gravitational
radiation.   Excellent summaries of these are given in Refs. 66 and 14.   We will
briefly discuss some of these.   Stellar locations for a few examples of each type
are listed in Tables VI-1-3.   Some are also plotted in Figure III-10.

Source information is from a variety of catalogs and is generally in right
ascension α, and declination δ relative to earth mean equator and equinox of 1900
or 1950.   We have adjusted all these to A.D. 2000 for uniformity. This is essentially

the system E^G in Section III-3.

To determine focal mission performance we need the equivalent angles α' and δ' relative to the ecliptic; i.e., the system E^S in Section III-3. Applying III-4 the connection between these is:

$$\begin{bmatrix} c\alpha'c\delta' \\ s\alpha'c\delta' \\ s\delta' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c\epsilon & s\epsilon \\ 0 & -s\epsilon & c\epsilon \end{bmatrix} \begin{bmatrix} c\alpha c\delta \\ s\alpha c\delta \\ s\delta \end{bmatrix} = \begin{bmatrix} c\alpha c\delta \\ s\epsilon s\delta + c\epsilon s\alpha c\delta \\ c\epsilon s\delta - s\epsilon s\alpha c\delta \end{bmatrix} \quad (1)$$

Since $|\delta, \delta'| \leq \pi/2$, the ambiguities are easily resolved.

One well studied source type is binary stars. The general theory of these is sketched in MTW[66], and worked out in detail for stars in Keplerian orbits by Peters and Mathews[70]. Post-Newtonian corrections have been calculated by Wagoner and Will[91]. All these are based on general relativity; but Will[95] has examined the same problem within several competing theories of gravity. These theories differ widely on what binary systems should emit, and on what our detectors should observe, which accounts for much of the interest in binaries.

The main difficulty with binaries as focal mission candidates is their low radiation frequency (twice their orbital frequency, in general relativity). This leads to such long wavelengths that diffraction (ignored in Chapter II) may wash out solar focussing. (See also Section 3 below.) However, on the chance that future studies do not confirm this effect, several of the most promising binaries are noted in Table VI-1. These are taken from Refs. (66), (14), and a recent compilation by Douglass[28]. The values of the radiation frequency, flux density, and strain amplitude are all by Douglass, who based his calculations on the theory of Peters and Mathews[70].

Table VI-1. Binary Sources

Source	Equatorial Position		Ecliptic Position		Wave Period sec	Flux Density $10^{-15}W/m^2$	Strain Amplitude x 10^{-21}
	α-deg	δ-deg	α'-deg	δ'-deg			
i Boo	225.947	47.654	197.778	60.537	11569	68	5.1
V Pup	119.559	-49.245	145.701	-67.055	62834	1.9	4.6
WZ Sge	301.892	17.695	309.015	36.914	2440	37	.8
YY Eri	33.075	-10.261	27.094	-22.151	13900	4.4	1.5
SW Lac	343.425	37.914	2.872	40.791	13900	3.2	1.3
VW Cep	309.794	75.396	52.085	74.775	12023	23	3.0
μ' Sco	252.965	-38.05	256.154	-15.426	62480	38	21
AM CVn	188.721	37.648	170.368	37.457	526	240	.5
SS Cyg	325.675	43.583	350.461	52.655	11904	20	3.0

Note: Positions referred to equinox of A.D. 2000. Obliquity = 23.4393 deg.

For our purposes, the table shows that the strongest sources from an energy standpoint are, in order, AM CVn, i Boo, μ' Sco, and WZ Sge; but for detectors directly sensitive to strain, the best sources are μ' Sco, i Boo, and V Pup, in that order. On the other hand, the best candidates from a rocket performance standpoint are μ' Sco, YY Eri, WZ Sge, and AM CVn. Unfortunately, none of their foci are reachable by direct chemical ascent.

A second interesting source type is pulsars. Accepting their current interpretation as spinning neutron stars, gravitational radiation at twice the spin frequency is expected if they possess unequal equatorial moments of inertia. Compelling evidence does not now exist that there could be sufficient deformation to produce detectable radiation; nevertheless, experimenters are presently exploring ways to detect them, based on their precisely known frequencies.

For one such attempt, see Sections V-3-5.

Some of the more interesting pulsars are listed in Table VI-2. Of these the Crab Nebula (M1) and Vela X are included because they have high spin rates, and thus are likely to be the strongest emitters. CP 0950 + 08 is included because it is relatively close (~ 60 pc). It would produce flux here comparable to, say, M1 with 800 times less output. Finally, although fairly distant, PSR 1913 + 16 is doubly interesting because it is a pulsar in a binary system. Neither its binary (wave period = 13,954 sec) nor its spin radiation is likely to be very strong, as seen here; but astrophysical interest in this system is currently very high, and detection of either one could advance both pulsar and relativity theory. The interesting complication here, as pointed out by Wagoner[90], is that the binary radiation may lead to observable orbit decay. If decay is observed, and if mass transfer can be definitely excluded as the cause, an indirect observation of gravitational radiation could be inferred. For recent estimates of M1 and Vela X emissions, see Ref.100.

Table VI-2. Pulsars

Pulsar	Equatorial Position		Ecliptic Position		Wave Frequency Hz
	α-deg	δ-deg	α'-deg	δ'-deg	
Crab (M1)	83.626	22.018	84.091	-1.291	60.427
Vela X	128.417	-45.788	153.487	-61.039	22.419
CP 0950 + 08	148.29	7.928	147.708	-4.619	7.903
PSR 1913 + 16	288.868	16.094	293.231	38.024	33.881

As for reaching pulsar foci, both M1 and CP 0950 + 08 have very low δ'. Moreover, as seen in Fig. III-10, both are inside the optimal shuttle orbit band. This puts them well within the capability of the chemically propelled

spacecraft of Chapter III. On this ground, these two pulsars must be regarded as prime candidates for a focal mission.

Another potential source of gravitational radiation is black holes. While fully convincing evidence for these does not exist at this writing, some are thought to be located at the centers of galaxies or globular clusters, where they might consume whole stars fairly frequently. This would presumably cause bursts of gravitational radiation, with the energy centered in the kilo-hertz region. For discussions of this see Ref. 66, and more recently Refs. 88 and 87. We mention them here because detection of specific sources is easier at their foci, and because the angular resolution achievable there might contribute strongly to our understanding of these objects. A few plausible places to look for such sources are given in Table VI-3 below.

Table VI-3. Burst Sources

Source	Equatorial Position		Ecliptic Position	
	α-deg	δ-deg	α'-deg	δ'-deg
Galactic Center	266.543	-28.954	266.961	-5.551
M87	187.706	12.943	181.827	14.921
M82	149.019	69.697	118.988	52.132
NGC 5195	202.499	47.259	175.091	50.999
NGC 1851	78.513	-40.026	70.585	-62.693
NGC 6624	260.925	-30.431	262.12	-7.24

Our galactic center is listed for several reasons. First, in Weber's original claims to have detected gravitational radiation (see V-2), he cited the galactic center as consistent with his weakly directional results. Since

then there has been much speculation about the presence of a black hole there, partly supported by the known strong infrared and radio emissions. Another reason for this interest is that our galactic center is surely the closest place we could plausibly look for a black hole of the order of 10^7 solar masses or more.

This idea was soon extended to more distant objects. For instance, Thorne and Braginsky[88] have suggested that any galaxy with an active nucleus (quasars, Seyferts, strong radio emitters) might contain massive black holes. The strongest evidence for this comes from the recent optical measurements of the giant elliptical galaxy M87 by Sargent et al.[76] and Young et al.[99]. They find that their central luminosity and velocity dispersion data are consistent with a central tiny non-luminous mass of about 5×10^9 solar masses. Such a hole would be larger than the solar system, and might consume stars fairly often. In support of this it may be noted that M87 is a strong radio and X-ray emitter, and it possesses a luminous unidirectional nuclear jet. As a member of the nearby Virgo cluster of galaxies, it may be the best prospect for a gravitational radiation source. However, radiation from stellar motions about the hole may be dominant over that from the consumption of stars by the hole, since[70] the former increases with hole mass, while[66] the latter is inversely proportional to hole mass. Of course, as this kind of measurement is extended to other galaxies, more black hole candidates are likely to appear. Some other possibilities include the "exploding" galaxy M82, and NGC 3077, which are highly irregular companions of the nearby M81 galaxy, and NGC 5195, an equally irregular companion of the "Whirlpool", M51.

For our last sources, we note that recent measurements of globular

clusters have shown that a few emit strong bursts of X-rays. This has been taken by Bahcall and others[6],[7],[8] as evidence for central black holes. The bursts are suggested to originate from clumps of matter entering an accretion disk, and from compact stars orbiting through the disk. The best candidates for such structures are given by these authors as NGC 1851 and NGC 6624. The chance for detecting gravitational radiation from these appears to lie in the direct consumption of stars (which must be very rare) or to large irregularities in the rotating accretion disk.

Finally we need to look at rocket performance. Table VI-3 shows that only the galactic center and NGC 6624 are accessible by chemical rockets, although they are slightly outside the optimal shuttle orbit band. While M87 has $\delta' = 15$ deg and would thus appear to have some chance for an all chemical mission, it is well outside the shuttle band; so without a gravity assist it does not look very feasible. However, at this low declination, electrical propulsion offers payloads in excess of 9 Mg - an attractive possibility.

3. Lens Problems

There are several difficulties with the solar lens which are not considered in Chapter II. Perhaps the worst of these is diffraction. Our analysis corresponds to the geometric optics assumption for electromagnetic radiation. The latter breaks down when the wavelength of the radiation approaches the lens system's effective aperture. Applying this analogy to the solar gravitational lens, we could conclude from Figure A-1 that the effective aperture is ~0.15 solar radii, or ~10^8 m. This corresponds to a frequency of ~3 Hz. If this is correct, then radiation from binary systems is not significantly focussed by the sun, and only burst sources and the most rapid pulsars would give anything like the picture in Figure A-3.

Another diffraction possibility is that the effective aperture would correspond to the Schwartzchild radius of the sun (1477 m); but this seems unlikely without a massive collapsed core. Yet another view is that the effective aperture should be taken as the radius of space curvature near the center of the sun. Taking a rough density radius of 0.1 solar radii, Figure A-1 indicates $W\sim60$, or $\rho\sim85,000$ kg/m^3. The radius of curvature is then (Ref. 66, Eq. 17.7):

$$r_c = \frac{c}{2\sqrt{\pi\rho G}} = \frac{3 \times 10^8}{2[\pi(85,000)(6.673 \times 10^{-11})]^{\frac{1}{2}}} = 3.56 \times 10^{10} \text{ m}$$

corresponding to a frequency of ~0.01 Hz. Better, but still not very hopeful for binary systems. This whole area needs attention by physicists.

A different lens problem is interference. This can occur behind the caustic where three rays intersect at each point. At sufficiently high frequency, differences in coordinate time along the three rays can lead to destructive interference. To get an estimate of this effect we can start from the gravitational time delay for radio signals passing by the sun, as obtained in Ref. 66, Eq. 40.14. For one-way travel, as applied to our problem, this delay is approximately

$$\Delta\tau(R_0) = \tau_1 \ln \frac{L(0)}{R_0} \tag{2}$$

where $\tau_1 = 4Gm_s/c^3 = 1.97 \times 10^{-5}$ sec. Applying this to a ray passing through the sun, and assuming that the instantaneous effect is due solely to the mass inside the radius R, (2) becomes:

$$\Delta\tau(R_0) = \tau_1 M(R_0)\ln \frac{L(0)}{R_0} + \tau_1 \int_{R = R_0}^{R = 1} \ln\left[\frac{L(0)}{R}\right] dM$$

$$= \tau_1 \left[\ln L(0) + \int_{R_0}^{1} \frac{M}{R} dR\right] \tag{3}$$

To this we must add the geometric delay due to the path being of approximate length $\sqrt{L(0)^2 + R_0{}^2}$ instead of $L(0)$. Thus the difference between rays with perihelia R_0' and R_0 is

$$\delta\tau = T(R_0') - T(R_0) \tag{4}$$

where

$$T(R_0) \doteq \tau_2 R_0^2 - \tau_1 \int_0^{R_0} \frac{M}{R}\, dR \tag{5}$$

and $\tau_2 = r_s{}^2/[2\ell(0)c] = 2.1 \times 10^{-4}$ sec. This function has been computed for for the Ezer-Cameron model, using M/R from Table A-2 and the parabolic integrator of Appendix C. The results are shown in Table VI-4.

Table VI-4. Interference Function

R_0	0	.05	0.140	0.154	0.2	0.295	0.399	0.495	1.0
$T(R_0)$ – μsec	0	0.449	1.69	3.42	5.47	11.84	22.85	37.16	181.9

For the region near $L(0)$ we are generally dealing with rays for which $R_0 < 0.2$. Thus path differences will not exceed about 5 μsec, and interference effects should not be significant below about 20 kHz.

Still another problem with the lens is that it is not perfectly spherically symmetric; i.e., the assumption $W = W(R)$ in Chapter II is not exact. One cause of this is solar rotation, which must cause some oblateness. Another possibility is convection cells, which, if they exist, would cause tesseral as well as zonal variations in density. Also sunspots and "M" regions are certainly irregular, although they may be superficial. Whatever their cause, aspheric density variations will distort the axial symmetry of the focal region, as in Figure A-3. This is both annoying, in that we cannot rely on axial symmetry, and hopeful, in that a careful mapping of the shape of the focal region might yield much information on solar structure.

One other effect can smear the focal picture - image size, although this is not strictly a lens problem. For a source of physical dimension w, at a distance D, the image size is

$$y = w\ell(0)/D \tag{6}$$

For instance, the binary i Boo has a projected separation of w ~ 1 Gm. At a distance $D = 3.9 \times 10^{17}$ m, this gives y ~ 9 km. Thus, image size is only a minor effect for binaries, and will be negligible for pulsars and other collapsed objects. Of course when the mission objective is to study the source, rather than the sun, image size is of great importance. This point is further discussed in Section 5.

4. Navigation

An engineering problem of considerable difficulty, not so far addressed, is navigation. Hopefully, the author, a candidate specializing in Guidance and Control, will be forgiven this transgression. The problem divides into two parts - the absolute location of the axis, and the navigation of the space-craft along the axis beyond the turn point. We will look at both of these and suggest some approaches.

What makes focal mission navigation unique amongst deep space probes is the lack of terminal guidance. Although all the possible sources discussed in Section 2 are detectable optically, they are necessarily eclipsed by the sun near the axis. Thus, indirect methods of navigation will be needed. The overall required accuracy of this process is set by the radial dimensions of the focal pattern. Taking the Ezer-Cameron solar model as a rough guide, and assuming that we would desire to find the gain = 1000 contour in the forelobe of Figure A-3, we find that the combined error must be held to $\lesssim$ 300 km.

Of this total, we might allot, say, 100 km to the absolute location of the source. At a distance of 3.5 Tm this yields a source stellar position accuracy of 3×10^{-8} rad or about 0.006 arc sec. Astrometry at this level is not elementary. However, for radio sources, very long baseline interferometry may be helpful.

Another major source of uncertainty is the solar model. The actual sun might yield a substantially smaller gain = 1000 contour than the Ezer-Cameron model. Also, solar asphericity could significantly warp the axial symmetry we have assumed, thus flattening the focal pattern somewhat. Finally, general relativity may not be correct, thus invalidating the null geodesic analysis of Chapter II. (This is distinct from the diffraction problem discussed in Section 3.) We will allot 100 km to these combined problems, realizing that any such number is hard to defend.

The remaining 100 km can be reserved as a navigational tolerance, the feasibility of which we will now examine. To begin with, axial navigation is not at all critical, so we are faced with only a two dimensional problem. Moreover, small motions orthogonal to the axis are unaffected by the solar field; so even these two dimensions are uncoupled, except, possibly, for correlations in the noise, which, for simplicity, we will ignore. Thus, we need to look at the case of controlling one-dimensional, field free motion, with perturbing acceleration, and relying on a noisy measurement of position.

Physically we will suppose that the control is by continuous low level ion propulsion. The perturbing acceleration may arise from unknown bodies in the solar system, uncertainties in the planetary masses, solar pressure, radiation from the spacecraft's power source, evaporation or outgassing of propellants and surfaces, and errors in the application of the controls. The

last is probably the most serious for a spacecraft beyond 2 Tm, and which has already spent several years in space. As for the measurement, we will temporarily assume that the position is continuously monitored by the NASA Deep Space Network, using earth based antennas.

The arrangement just discussed can be described by $\ddot{x} = u + w$, where x is the position error, and u and w are the control and perturbing accelerations respectively. In order to apply existing theory, we will recast this in state variable form:

$$\dot{x} = Fx + Gu + W \tag{7}$$

where

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} ; \quad F = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} ; \quad G = \begin{bmatrix} 0 \\ 1 \end{bmatrix} ; \quad W = \begin{bmatrix} 0 \\ w \end{bmatrix} \tag{8}$$

Again for simplicity we will take the perturbation w as a stationary random process with zero mean and autocorrelation

$$E[w(t)w(t')] = q\delta(t-t') \tag{9}$$

Also, the measurement

$$z = Hx + v ; \quad H = [1,0] \tag{10}$$

is assumed to contain a noise v with similar properties:

$$E[v(t)v(t')] = r\delta(t-t') \tag{11}$$

In reality, both noises are time correlated, but a full treatment is much harder analytically, and is hardly justified by the quality of the available statistical data. However, we do have to relate q and r to published numbers; so for this we will assume an exponential autocorrelation:

$$q = 2\tau_w \sigma_w^2 ; \quad r = 2\tau_v \sigma_v^2 \tag{12}$$

where τ_w and τ_v are the correlation times, and σ_w and σ_v are the standard deviations. For a fuller discussion of this see Ref. 17, Section 11.4.

According to the certainty - equivalence principle (Ref. 17, Sections 14.3-5) the optimal controller for the above system starts with a state estimator:

$$\dot{\hat{x}} = F\hat{x} + Gu + K(z - H\hat{x}) \tag{13}$$

in which the feedback gains K are given by

$$K = \frac{1}{r} PH^T = \frac{1}{r} \left[P_{11}, P_{12} \right]^T \tag{14}$$

and in which

$$\dot{P} = FP + PF^T - \frac{1}{r} KK^T + Q \tag{15}$$

In this

$$Q = \text{diag} \left[0, q \right] \tag{16}$$

and we have assumed w and v to be uncorrelated (another somewhat shaky assumption, as we will see). The matrix P is the covariance of the estimator error:

$$P = E\left[(\hat{x} - x)(\hat{x} - x)^T \right] \tag{17}$$

Since there is no explicit time dependence in our problem, and as there is plenty of time after the turn point for the system to settle, we can take K as constant and look for a steady state solution ($\dot{P} = 0$). The above algebraic system may be solved for P. Since the variances P_{11} and P_{22} cannot be negative, we get:

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{12} & P_{22} \end{bmatrix} = \begin{bmatrix} \sqrt{2r\sqrt{qr}} & \sqrt{qr} \\ \sqrt{qr} & \sqrt{2q\sqrt{qr}} \end{bmatrix} \tag{18}$$

from which the gains are:

$$K = \left[\left(\frac{4q}{r} \right)^{\frac{1}{4}}, \left(\frac{q}{r} \right)^{\frac{1}{2}} \right]^T \tag{19}$$

With the state estimate $\hat{x}$ available we now turn to the controller:

$$u = -C\hat{x} \tag{20}$$

With the idea that we wish to minimize $E(x_1^2)$ while not letting the propulsion requirement get too big, we form the quadratic performance index:

$$J = \frac{1}{2} E\left[\int (ax_1^2 + bu^2)dt\right] \quad \text{(21)}$$

where the integral is over some long time. The optimal gains are then:

$$C = \frac{1}{b} G^T S = \frac{1}{b} \left[S_{12}, S_{22}\right] \quad \text{(22)}$$

and

$$\dot{S} = \frac{1}{b} C^T C - SF - F^T S - A \quad \text{(23)}$$

where

$$A = \text{diag } [a,0] \quad \text{(24)}$$

Once again we look for a stationary solution ($\dot{S} = 0$). For a positive semi-definite S we get:

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix} = \begin{bmatrix} \sqrt{2a\sqrt{ab}} & \sqrt{ab} \\ \sqrt{ab} & \sqrt{2b\sqrt{ab}} \end{bmatrix} \quad \text{(25)}$$

from which:

$$C = \left[\left(\frac{a}{b}\right)^{\frac{1}{2}}, \left(\frac{4a}{b}\right)^{\frac{1}{4}}\right] \quad \text{(26)}$$

To evaluate the performance of the system, we need the covariance of the estimated state $\hat{X}$. This can be found from:

$$L\hat{X} + \hat{X}L^T = rKK^T \quad \text{(27)}$$

where

$$L = GC - F = \begin{bmatrix} 0 & -1 \\ C_1 & C_2 \end{bmatrix} \quad \text{(28)}$$

The solution of this system is:

$$\hat{X} = \begin{bmatrix} \hat{x}_{11} & \hat{x}_{12} \\ \hat{x}_{12} & \hat{x}_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{C_1}\left(\frac{q}{2C_2} + \frac{3}{2}C_{12}P_{12} + P_{22}\right) & -P_{12} \\ -P_{12} & \frac{1}{2}\left(\frac{q}{C_2} + C_2 P_{12}\right) \end{bmatrix} \tag{29}$$

From this we can form a more interesting quantity - the covariance of the actual state:

$$X = \hat{X} + P = \text{diag}\left[P_{11} + \frac{1}{C_1}\left(\frac{q}{2C_2} + \frac{3}{2}C_2 P_{12} + P_{22}\right), \ P_{22} + \frac{1}{2}\left(\frac{q}{C_2} + C_2 P_{12}\right)\right] \tag{30}$$

Also of interest is the control variance:

$$E(u^2) = c\hat{X}c^T = \frac{1}{2}C_1\left(\frac{3q}{C_2} + C_2 P_{12} + 2P_{22}\right) \tag{31}$$

We are now ready to introduce some numbers. A good way to get trial design numbers for the control loop is to invoke "Bryson's Rule" - choose weights a and b according to the maximum tolerable state and control excursions. Since we have allotted 100 km for navigation:

$$a = (10^5)^{-2} = 10^{-10} \ m^{-2}$$

As for control, the average acceleration for a 7 year ion drive trajectory is around $2 \times 10^{-5} \ m/sec^2$, so if we allow 5% of this for control we get:

$$b = (10^{-6})^{-2} = 10^{12} \ sec^4/m^2$$

From (26) this immediately gives us trial control gains:

$$C_1 = \left(\frac{10^{-10}}{10^{12}}\right)^{\frac{1}{2}} = 10^{-11} \ sec^{-2}$$

$$C_2 = \left[\frac{(4)(10^{-10})}{10^{12}}\right]^{\frac{1}{4}} = 4.472 \times 10^{-6} \ sec^{-1}$$

It is important to look at the timing associated with these gains. We can find this from the system's closed loop eigenvalues:

$$|\lambda I + L| = \begin{vmatrix} \lambda & -1 \\ c_1 & c_2 + \lambda \end{vmatrix} = 0 \qquad (32)$$

whose roots in this case are:

$$\lambda = \frac{1}{2} c_2(-1 \pm i) = 2.236 \times 10^{-6}(-1 \pm i) \text{ sec}^{-1}$$

Thus damping times run about 4.5×10^5 sec or 5.2 days.

One reason for treating the controls first is that they may be the most important contributor to the perturbation w. There is no way, at this point, to get an accurate estimate of σ_w; however 10% of the maximum control looks very conservative. Thus, $\sigma_w = 10^{-7}$ m/sec^2. We will return to this point. As for the correlation time, the above damping time is the obvious choice. Thus, from (12):

$$q = (2)(4.5 \times 10^5)(10^{-7})^2 = 9 \times 10^{-9} \text{ m}^2/\text{sec}^3$$

Next we will look at the measurement noise. The current and projected performance of the NASA Deep Space Network is reviewed by Melbourne and Curkendall in Ref. 62. They find that the present system, used at X-band in the differential range mode between distant ground stations, is capable of measuring angles to an accuracy of 5×10^{-8} to 2.5×10^{-7} rad., depending on the tracking time allowed, and the effort they are willing to spend to keep the system calibrated. This arrangement is especially attractive because it does not require an uplink, and because it does not have unusual difficulties at low equatorial declinations. Considering that we will need nearly continuous measurements for a period of several years, we cannot expect a sustained maximum effort at calibration; so we will adopt the conservative bound. At the Ezer Cameron focal range of 3.5 Tm this gives $\sigma_v = 875$ km.

As for the correlation time, the same authors assume a spacecraft power of 1 watt, radiated from an aperture of 3.7 m at 50% efficiency, and at a distance of 3 Tm. With the present 64 m ground antennas at temperatures of 30 K they find that adequate post-detection signal-to-noise ratio is achieved with 8 sec of integration time. The difficulty with this is that we cannot reasonably expect to tie up the main tracking facilities for several years. However, these authors also point out that much less expensive 10 m antennas at temperatures of 100 K would only require 1000 sec of integration time for the same result. For our purpose we will assume that facilities of this quality exist.

As for our spacecraft, the antenna seems about right, but we are at a slightly greater distance. On the other hand, we could easily raise the radiated power. Thus we can regard the integration time of 1000 sec as somewhat conservative, and adopt it as the correlation time. Putting these numbers in (12):

$$r = (2)(1000)(8.75 \times 10^5)^2 = 1.531 \times 10^{15} \text{ m}^2\text{-sec}$$

With these, and (18), we compute the estimator error covariance:

$$P_{11} = 3.372 \times 10^9 \text{m}^2; \quad P_{12} = 3712 \text{ m}^2/\text{sec}; \quad P_{22} = 8.174 \times 10^{-3} \text{ m}^2/\text{sec}^2$$

and from (14) the estimator gains:

$$K_1 = 2.202 \times 10^{-6} \text{ sec}^{-1} ; \quad K_2 = 2.424 \times 10^{-12} \text{ sec}^{-2}$$

The estimator dynamics can be found in the same way as in the control loop. Here the characteristic equation is:

$$|\lambda I + L + KH| = \begin{vmatrix} \lambda + K_1 & -1 \\ C_1 + K_2 & C_2 + \lambda \end{vmatrix} = 0 \tag{33}$$

whose roots are:

$$\lambda = \frac{1}{2}(C_2 + K_1)(-1 \pm i) = 3.337 \times 10^{-6}(-1 \pm i) \text{ sec}^{-1}$$

Thus, the estimator's settling time runs around 3×10^5 sec, or 3.5 days, which is a little faster than the control.

The other quantities of interest are the covariance of the estimate from (29):

$$\hat{X}_{11} = 3.408 \times 10^9 \text{ m}^2; \quad \hat{X}_{12} = -3712 \text{ m}^2/\text{sec}; \quad \hat{X}_{22} = 9.307 \times 10^{-3} \text{ m}^2/\text{sec}^2$$

the state covariance from (30):

$$X_{11} = 6.78 \times 10^9 \text{ m}^2; \quad X_{12} = 0; \quad X_{22} = 0.01748 \text{ m}^2/\text{sec}^2$$

and the control variance from (31):

$$E(u^2) = 1.949 \times 10^{-13} \text{ m}^2/\text{sec}^4$$

The most important measures of performance are the standard deviations of the position and control:

$$\sigma_x = \sqrt{X_{11}} = 82.34 \text{ km}; \quad \sigma_u = \sqrt{E(u^2)} = 4.415 \times 10^{-7} \text{ m/sec}^2$$

Assuming that these numbers describe normal distributions with zero mean, we find that the position error exceeds our requirement of 100 km 22.46% of the time, while the control exceeds the 10^{-6} m/sec limit 2.35% of the time.

Two points stand out from this analysis. First, the percentages indicate that the performance weight ratio a/b has been chosen too low. Accordingly, it was decided to raise this by a factor of 4 and try again. At the same time it was noted that the comparison between σ_u and the process noise σ_w shows that we have been overly conservative in the choice of the latter. In the second trial σ_w was reduced to $\sigma_u/10$. Both changes reduced q since the tighter control dropped the control damping time. The results of the new calculations are shown side by side with the older ones in Table VI-5.

TABLE VI-5. NAVIGATION DESIGN PARAMETERS

PARAMETER	DESCRIPTION	UNITS	VALUE		
a/b	Weight ratio	sec^{-4}	10^{-22}	4×10^{-22}	1.6×10^{-21}
C_1	Control position gain	sec^{-2}	10^{-11}	2×10^{-11}	4×10^{-11}
C_2	Control rate gain	sec^{-1}	4.472×10^{-6}	6.325×10^{-6}	8.944×10^{-6}
	Control damping time	days	5.2	3.7	2.3
σ_w	Process noise σ	m/sec^2	10^{-7}	4.415×10^{-8}	4.4×10^{-8}
q	Process noise autocorrelation	m^2/sec^3	9×10^{-9}	1.233×10^{-9}	8.658×10^{-10}
σ_v	Measurement noise σ	km	875	875	875
r	Measurement noise autocorrelation	m^2-sec	1.531×10^{15}	1.531×10^{15}	1.531×10^{15}
P_{11}	Estimator	m^2	3.372×10^9	2.051×10^9	1.878×10^9
P_{12}	error	m^2/sec	3712	1374	1151
P_{22}	covariance	m^2/sec^2	8.174×10^{-3}	1.841×10^{-3}	1.412×10^{-3}
K_1	Estimator rate gain	sec^{-1}	2.202×10^{-6}	1.34×10^{-6}	1.226×10^{-6}
K_2	Estimator acceleration gain	sec^{-2}	2.424×10^{-12}	8.973×10^{-13}	7.519×10^{-13}
	Estimator settling time	days	3.5	3.0	2.6
$\hat{x}_{11}$	Covariance	m^2	3.408×10^9	7.486×10^8	4.227×10^8
$\hat{x}_{12}$	of the	m^2/sec	-3712	-1374	-1151
$\hat{x}_{22}$	estimate	m^2/sec^2	9.307×10^{-3}	4.442×10^{-3}	5.198×10^{-3}
σ_x	Position σ	km	82.34	52.91	47.96
σ_u	Control σ	m/sec^2	4.415×10^{-7}	3.599×10^{-7}	5.179×10^{-7}
	Position error > 100 km	%	22.46	11.75	3.71
	Control > 10^{-6} m/sec^2	%	2.35	0.55	5.35

In comparing performance, note the graftifying drop of σ_x from 82.34 km to 52.91 km. Also, σ_w and σ_u now appear more consistent. However, the percentages of time outside the limits are still somewhat disparate. Thus, one more trial was made with a/b raised another factor of 4. These results are shown in the last column of Table VI-5. This time the control usage σ_u is up significantly without much further improvement in σ_x. The "best" design choice probably lies somewhere between the last two columns. With the 100 km navigational tolerance exceeded only a few percent of the time, and settling times of ~3 days, it does not seem plausible that a feature as large as the forelobe of the focal pattern (Figure A-3) could be missed.

The long estimator settling times are actually somewhat advantageous in that tracking data lapses of up to several hours should not seriously degrade performance. This conclusion is reinforced by observing that typical cross axis speeds are about $\sqrt{X_{22}}$ ~0.08 m/sec, so that the displacement during one day is typically only ~7 km. We conclude that while facilities at two (or preferably three) stations must devote nearly full time to this mission, combined two station coverage during 70-80% of the time is probably adequate.

Several ways exist for further improving performance. Ground antenna size could be increased, or receivers improved, or the number of stations increased, or some combination of these. All are likely to be expensive. Similarly, the spacecraft's antenna size or radiated power could be increased. This may be feasible on electrically propelled spacecraft but not on the much smaller chemical craft. A related possibility is that r be reduced impulsively, either by temporary increases in radiated power to decrease τ_v, or by making quick fixes with the main Deep Space Network facilities every day or so to reduce σ_v. The present steady state analysis will not accommodate a variable r; but these measures would obviously help.

Another approach is to reduce q. This could be done by carrying an on-board accelerometer. Perhaps the simplest way to incorporate this measurement is to replace u in the estimator structure (13) by the measured acceleration. This replaces the actual process noise w by the accelerometer error. Present technology in rebalance accelerometers is $\sigma \sim 10^{-8}$ m/sec, as exemplified in the French CACTUS instrument. Values of $\sigma \lesssim 10^{-9}$ m/sec^2 appear to be achievable in drag free accelerometers[82]. In this case, drag free periods of 100 sec or more with free motion of $\pm$ 1 mm requires $\sigma_u \leq$ 2 x 10^{-7} m/sec^2, which appears feasible from the above analysis. For either instrument, the correlation time should be very much less than the $\tau_w \sim$ 3 days we have been using.

While these numbers would appear to reduce q dramatically, there is one difficulty - gravitational perturbations, due to uncertain planetary masses or unknown solar system bodies, are not measured by accelerometers. Suppose our tolerance for this is 10^{-9} m/sec^2. Then an unknown mass M must be at least a distance R away, given by:

$$M/R^2 = 10^{-9}/G = 14.99 \text{ kg/m}^2$$

For instance, Saturn has a mass of about 5.69 x 10^{26} kg; so a 1% error (really a 1% error in GM_{SAT}) in its determination requires us to avoid Saturn by at least 616 Gm, which is no problem beyond the turn point. As for Neptune at 1.03 x 10^{26} kg, the corresponding distance for a 2% error is 370 Gm. Uranus is slightly smaller than this, and less troublesome. Most missions would not come nearly this close to an outer planet; but if it were to occur, a year's delay of the mission would very likely remove the problem.

The other possibility is an unknown solar system body. To get a bound on this we can suppose that our chance of encountering an unknown body of, say, 500 km diameter is remote. Still if it were spherical, and had an earth-like density of 5000 kg/m^3, it would have M = 3.27 x 10^{20} kg. It would give a disturbance of 10^{-9} m/sec^2 at a distance of 4.67 Gm. Since only one asteroid (Ceres) is known to be larger or more massive than this, trouble from this source seems highly unlikely. If, in spite of all this, we have such an encounter, it would affect only a small fraction of the trajectory. Moreover, such a perturbation might reveal itself by its unique signature in the tracking residuals.

In conclusion, it appears that navigation to 100 km accuracy is feasible with nearly continuous Deep Space Network tracking, using modest 10 m antennas and 100 K receivers. Better performance is achievable if the non-gravitational disturbances are measured with an accelerometer at the 10^{-8} to 10^{-9} m/sec^2 accuracy level. However, to reach this result, drastic simplifications have been employed, and a much more elaborate analysis is called for. Also, the possibilities for replacing or augmenting Deep Space Network tracking, such as celestial navigation using the planets, or interplanetary navigational beacons, have not been considered, but could play an important role.

5. Mission Optimization

A good deal has been said about optimal trajectories up to the turn point, but, save for navigation, little about the remaining trajectory phases. Here we will look at how the post turn point trajectory depends on the mission objectives. This is an extremely complex area, and a great deal of work will be needed to maximize the scientific yield of any given mission.

To expose the main ideas we will examine three types of missions, each with a single main objective, realizing that this is an over-simplification. These types are the first mission, solar missions, and astronomical missions.

The first mission is unique because we don't know the focal length. Because of this, we will take the measurement of the focal distance $\ell(0)$ as the primary mission objective. The main benefit of this is that later missions can, presumably, choose more distant turn points, thus yielding shorter mission times, or larger payloads, or both. Also, from (II-101), $\ell(0)$ is a direct measure of δ_0^-, which in turn (II-83), yields $\int_0^1 W(R)dR$. This would be of great value to solar physics in sorting out models of the core density of the sun.

The profile of a first mission is not hard to work out. The main events are:

A.   Preliminaries. Sources are detected somehow, and an appropriate spacecraft-detector combination is designed and built. Low source ecliptic declination is an important selection criterion.

B.   Transfer Orbit. A turn point radius $r_T \sim 2$ Tm is selected, as discussed earlier. This gives the payload mass as a function of the post-turn point speed in the chemical case, or as a function of this speed and the transfer time for ion propulsion.

C.   Search Phase. Since finding the focus is our principal objective, any excess payload capacity should be devoted to propulsion, in order to minimize the mission time (and thus cost). For any given suspected $\ell(0) = \ell_s$ we could work out the optimal orbital and propulsion parameters to yield the minimum overall time $t(\ell_s)$ from departure to ℓ_s. However, if $\ell(0) \neq \ell_s$, we will actually arrive at $\ell(0)$ at a time $t_0[\ell(o), \ell_s]$. In the case of the ion drive, we are powered up to ℓ_s, after which we are presumably out of propellant,

and must coast. Now suppose, based on existing knowledge of solar structure, that we are given an a priori estimate $p(\ell)$ of the likelihood density of finding the focus at ℓ. Then, assuming cost is linearly dependent on time, the optimal choice of ℓ_s is that which minimizes the performance index

$$J(\ell_s) = \int_{r_T}^{\infty} t_0(\ell,\ell_s) p(\ell) d\ell \qquad (34)$$

This problem should be straightforward for chemical power, but could be as difficult as Chapter IV for ion propulsion.

D. Final Phase. An examination of the gain diagram, Figure A-3, shows that we should begin to notice focal effects well before $\ell(0)$. Supposing that by the time the gain reaches 100 we would be certain that we had found the forelobe of the pattern, we would still have around 300 Gm to go. At an average speed of, say, 10 km/sec, it would take about another year to reach $\ell(0)$. This time could be put to good use by making small cross-axis maneuvers in order to locate the axis more accurately. Suppose we make a 1 Mm lateral shift over a 30 day perio This requires a velocity increment of 0.386 m/sec. If done by an ion thruster, the required acceleration is 3×10^{-7} m/sec^2, somewhat less than we assumed for navigation in the last section. Maneuvers of thi sort would continue, with diminishing amplitude, until $\ell(0)$ is passed With this, the main objective will have been accomplished; but additional maneuvers would be very desirable to get some focal details, especially the caustic location. The design of this maneuver sequenc to get an accurate determination of $\ell(0)$, while maximizing the chance of resolving other features, will be a challenging task.

We will now look at a solar mission. Here we will assume that $\ell(0)$ is known, and that the primary mission objective is a detailed map of solar structure. If, unlike Chapter II, we now admit that the sun may be at least weakly aspheric, even including tesseral variations, then an adequate description of the mass distribution would take the form of some kind of expansion. Mass multipole or spherical harmonic expansions of the density function are plausible candidates. In either case, the coefficients are constrained to yield the correct values of $\ell(0)$ and the solar mass.

Now suppose that a trial solar model is developed which is spherically symmetric, and consistent with the constraints. From it a focal gain picture similar to Figure A-3 can be computed. With further work, we can determine the partial derivatives of the gain at each point, relative to the general expansion coefficients. Then, in principle, by measuring the gain at a large number of points, we can invert the problem, and extract the corrections to the expansion coefficients. The result should be an improved, fully aspheric model of the sun. If the corrections are large, an iteration may be needed; but this does not necessarily require new measurements. The revised trial model could still be spherical.

In the case of long baseline detectors, the process is more complicated in that the predicted intensities and partial derivatives must be averaged over each baseline pair. The resulting loss of resolution would appear to give a distinct advantage to compact detectors for solar missions.

The profiles of the solar and first missions are quite different. First, the transfer orbit would be chosen to yield r_T around 200 Gm short of $\ell(0)$. Then, as a long search will not be needed, we could substantially reduce the

post turn point speed. Together, these would increase the possible payload, especially for chemically powered spacecraft. On this kind of trajectory, we can expect to detect focal effects immediately following the turn point with a chemical spacecraft, and even earlier with ion propulsion. We would then locate the axis with oscillatory maneuvers, similar to the first mission.

Following acquisition of the axis, the profile is not so clear. In any given trajectory through the focal area, we could compute the accuracy with which we could extract each of the solar model coefficients. Then, given a set of weights related to the scientific importance of each coefficient, we could devise a performance index for comparing trajectories. Different trajectories could be examined, and variations about the better ones could be explored; but whether a formal optimization, such as in Chapter IV, would be fruitful is unclear. One obvious problem is the gain discontinuity at the caustic, if this feature isn't washed out by diffraction.

It is the author's speculation that the most fruitful tour would be one that attempts to map the region a few kilometers either side of the caustic. A plausible way to do this would be to follow a spiral path on the caustic surface, with frequent surface crossings of a few kilometers depth. The dither could be expanded to allow for diffraction.

The larger payload capacity of electrically powered spacecraft opens the possibility of carrying several detectors, and sending them on separate trajectories following axis acquisition. One possibility is a set of planar trajectories evenly spaced around, and following the caustic. Better from the standpoint of cross-calibration of the detector sensitivities is a set of evenly spaced spirals, as mentioned above. Still another way would be to reverse the pitch of some of the spirals, so that the crossing points would give still better cross-calibrations. Other possibilities abound.

Three main motivations exist for detailed density measurements of the sun. First, an accurate determination of the symmetric reference model $W(R)$ could be immediately integrated to yield the pressure $P(R)$, and then, through equation of state arguments, the temperature. A great improvement in our understanding of solar structure is evident. Rather harder to measure is solar oblateness. The most obvious effect of this is that the circular cross section of the caustic would be replaced by an oval with symmetry axes determined by the solar equator. Measurements of the eccentricity of this oval at various distances beyond $\ell(0)$ could then measure the solar oblateness.

To get a rough estimate of the size of this effect, suppose the solar $J_2 \sim 10^{-5}$ (see Ref. 66 for a discussion of this). Presumably this would lead to an eccentricity of the oval of the same order (we have not shown this). Thus, in the region where the caustic radius is a few hundred kilometers, distortions of several meters are to be expected. While measurements between spacecraft to this accuracy should not be difficult, it remains to be seen whether the caustic can be resolved this sharply.

A successful oblateness measurement could settle the long standing argument about the precession of the perihelion of Mercury, and thus decide between the Einstein and Brans-Dicke relativity theories. (This may also be possible directly from focal structure.)

Finally we have some hope of discerning tesseral variations. Sources for these are possible convection zones, sun spots, and M regions. The latter are the source of certain terrestrial magnetic anomalies, varying with the synodic period of solar rotation, but not obviously connected with surface features. That tesseral variations are likely to be weak is only part of the problem.

They may also be expected to vary with the solar sidereal period, which, at least at the surface, is known to be latitude dependent. Still, any detected tesseral variations should be valuable to some aspect of solar structure.

For our last mission type, we will look at gravitational wave astronomy. Besides the obvious advantage of increasing the apparent source intensity, a focal mission permits very sharp angular discrimination. This helps both in eliminating nearby interfering sources, and in resolution of the source. As an example of the former, if two possible sources are separated visually by 1 arc second, their foci are separated by 17.3 Mm, if $\ell(0)$ = 3.5752 Tm (Ezer-Cameron).

As an example of the latter consider M87. If this galaxy really possesses a central black hole of 5 x 10^9 solar masses[76,99], then it has an event horizon of about 30 Tm diameter, if it is a pure Schwartzchild hole. At a distance of 13.6 Mpc and the same $\ell(0)$, the image of the event horizon is, from (6), a circle of 251 m diameter. An accretion disk would be somewhat larger. Even without diffraction, these features are probably unresolvable, except, possibly, right at $\ell(0)$. Stars in long elliptical orbits about the hole should radiate bursts of gravitational radiation near periapse. The effective diameter of the circle of the largest bursts would be perhaps an order of magnitude larger. At the other extreme, M87 has a "jet" extending about 19 arc seconds from the nucleus, including several very bright star-like objects. To examine the whole jet for gravitational activity would thus require a lateral excursion of about 330 Mm.

Another possibility is our galactic center. If gravitational radiation is somehow detected from that direction, a massive black hole will be suspected immediately. At, say, 10^7 solar masses (a wild guess), it would have an event horizon diameter of 59 Gm. At a distance of 10 Kpc (Ref. 1) its focal image

diameter is 1.85 m. Thus stellar encounters one or two orders of magnitude farther away may be resolvable, but capture events probably are not.

As for other possibilities, certain globular clusters, as discussed in Section 2, could possibly be resolved. Also, more galaxies with bright central luminosity cusps, similar to M87, are likely to be found. Galaxies with strong nuclear activity, such as Seyferts, may also be good candidates.

Quasars are a special case in that while they might radiate strong gravitational waves, their star-like appearance suggests that their focal images would be too small to resolve. Thus a mission to a single quasar focus would not appear to be fruitful. However, a mission involving a quasar and another nearby source, interesting by itself, may be valuable. Such a case might be the "exploding" galaxy M82, with its bevy of "nearby" quasars, more or less on a line stretching about 9 arc minutes from the center of M82. Whether they are really nearby, as suggested by Arp (Ref. 5), or are remote, as others believe, the true source(s) of possible gravitational radiation from that direction probably couldn't be decided without a focal mission. A lateral excursion of 8.3 Gm would be needed for this exploration.

That no mention of pulsars has been made here needs some comment. A neutron star of a few kilometers diameter might possibly possess an active accretion disk of 100 km diameter (doubtful), if it has nebular matter to feed on. At the relatively close distance of 500 pc, it would then have an image size of 23 mm. Thus pulsars are unresolvable, and there would appear to be little reason to stage an astronomical mission to a pulsar focus. On the other hand, a steady point source is ideal for a solar mission, so pulsar foci may be interesting after all.

In planning an astronomical mission, two points stand out from the above discussion - interest depends on the expectation of spatial structure, and interesting sources so defined are all ,expected to be random burst sources. The latter at least is subject to verification before actually planning a mission. Both these points complicate the process of focal acquisition as described for the first and solar missions. In fact, it is difficult to escape the conclusion that multiple detectors on more or less parallel trajectories would be needed, even supposing the source has a known history of fairly frequent bursts. Also, burst sources tend to concentrate their energy in the higher frequencies, which argues in favor of compact detectors - a necessity in any case, if fine source details are to be examined. Accepting this argument, we will assume that a fleet of compact detectors successfully acquires at least one component of the source. We can also conclude, regretfully, that astronomical missions are probably beyond the capacity of chemically powered spacecraft.

Because of the statistical nature of burst sources, their resolution imposes a unique requirement - that the fleet parks near $\ell(0)$. To hover in the solar field at $\ell(0)$ requires a steady acceleration:

$$a = \frac{F}{m} = \frac{\mu}{r^2} = \frac{1.3271 \times 10^{20}}{(3.5752 \times 10^{12})^2} = 1.038 \times 10^{-5} \text{ m/sec}^2$$

which is well within the capacity we have previously assumed. We conclude that for astronomical missions, ion thrusters would be needed more or less continuously from departure to the end of the mission. Whether they would have to be turned off during data periods, to avoid acoustic excitation of the detectors, is unclear.

Lateral transfers to nearby foci must be considered. Suppose we permit up to 6 months for a transfer, including 3 months each of acceleration and deceleration. Then at a = 2 x 10^{-5} m/sec^2 we would be permitted lateral transfers up to 2.5 Gm. At $\ell(0)$ = 3.5752 Tm, this will cover 144 arc seconds of the sky. Several such jumps during a mission could probably be accommodated.

To assemble these ideas into a profile, we suppose that one source component is selected on the basis that it is the probable source of the previously detected radiation. Since we will be approaching hover conditions as we near $\ell(0)$, a turn point should be selected even closer than in solar missions, perhaps 20 Gm short of $\ell(0)$. The fleet would be separated just after the turn point, and the initial acquisition process would start. Joint fleet maneuvers would occur after each detected burst in an attempt to equalize the reception of the next burst. Helical trajectories could be used to cross-calibrate the detectors, as in solar missions.

Once acquisition is complete, the fleet would dwell at $\ell(0)$ long enough to establish the source structure, if any, and then jump to the next potential focus. The order in which the different foci are visited should be chosen to minimize the total jump time. Since jump time is proportional to $\sqrt{S}$, where S is the lateral distance between successive foci, the tour should be chosen to minimize the performance index J = $\sum_i \sqrt{S_i}$. For complex tours, this is an interesting variation of the traveling salesman problem.

Acquisition of the new axis after a jump differs from initial acquisition because of the small lateral dimensions of the focus. For this reason, jumping to a region a little beyond $\ell(0)$ may be indicated. If radiation is not detected at the expected location, a search pattern may be indicated, in which the fleet is spread out to cover more space. The form of such a search is another interesting optimization problem. If no radiation is found by completion of the search raster, the fleet could depart directly for the next

C-4

focus. Alternatively, detection at any time would be followed by acquisition more or less as above, and concentration of the fleet at $\ell(0)$ for study. The planning of astronomical missions will obviously be very complex, but guided by the simple rule of allowing adequate times for jumps, searches, acquisitions, and dwells.

APPENDIX A

PUBLISHED SOLAR MODELS

1. Introduction

In planning missions to the solar foci we will need prior estimates of their radii. These can be computed from any given solar model; however the models are not directly verifiable, and thus contain errors which are difficult to bound. The problem is recursive in that a measurement of the solar density function is one of the mission objectives. Still, a number of published models exist, and some of them are examined here to see what focal length and caustic properties they predict. No opinion is offered as to the absolute or relative quality of these models, and the presence of some models here, and not others, should not be regarded as implied approval.

Of the five models selected, the one by Ezer and Cameron[39] was chosen because it was used in the earlier focal calculations of Cyranski and Lubkin[23]. For the same reason we have included Naur's second model (Naur II)[68]. The model Naur I was not treated by Cyranski and Lubkin; but it is included here, largely because it is also in Reference 68. To round these out, Martin Schwartzchild's* well known book on stellar structure[78] provides a rather detailed model. Finally, Allen's widely used reference[1] contains a model which is compiled and smoothed from several sources, and seemed appropriate here. The details of the models are given in nondimensional form in Tables A2-A6 and Figure A-1.

---

*Not the earlier Karl Schwartzchild for whom the line element is named.

2. Density, Mass, and Pressure

Two things are done to the basic tabular data of the references prior to Tables A2-A6. First, they are nondimensionalized according to the relations (II-32), where this is not already done. The different authors employ different values of r_s and m_s, and as these are fundamental parameters of the model, care is taken not to mix them up. Table A1 shows these values, as well as the derived values of γ and $\bar{\rho}$, the average density; the latter in g/cm^3, as these are the units of the references.

Table A1

Model		$10^{-30} m_s$ (kg)	r_s (Mm)	$10^6\gamma$	$\bar{\rho}$ (g/cm^3)
Ezer-Cameron	(A)	1.989	671	2.20066	1.57173
Naur I	(B)	1.989	695.5	2.12313	1.41141
Naur II	(B)	1.989	695.808	2.1222	1.409539
Schwartzchild		1.985	695.0	2.11963	1.41162
Allen 3rd Ed.		1.989	695.99	2.12088	1.40844
MTW Values		1.989	695.98	2.12167	1.4085

Notes: (A) m_s is not given, so the Allen-MTW value is used.

(B) Naur gives sufficient information to derive $\bar{\rho}$, but no other table entries. Thus the Allen-MTW value of m_s was assumed, and r_s and γ were then derived.

In all these sources radius and mass were already in the form R and M. Density was generally given in g/cm^3 (Naur listed $\rho/\rho(0)$), so W was readily computed using $\bar{\rho}$. Pressure was usually in dynes/cm^2 or $\log_{10}$ of that, so to

put it in the same nondimensional form as the density, it was necessary to divide by $\bar{\rho}c^2$, which yields P. In Naur's models, the pressure was not given; but a perfect gas is an ideal gas, so the pressure is given by:

$$R = \frac{n}{V} RT = \frac{\rho}{M_W} RT \qquad (A1)$$

where[42] R = 8.3143 joule/K-mol = universal gas constant, and M_W is the molecular weight. Naur lists the values M_W = 0.598 and 0.589 for the sun's mean molecular weight in his two models. Since the pressure is tiny compared to the density, a more elaborate model of M_W is hardly warranted. At any rate, this gave sufficient information to compute P from Naur's tables, using (A1).

The other problem with these tabular models is that they tended to be sparse. Even Schwartzchild's model is too coarse in the critical central region. To alleviate this, two forms of interpolation were used. First, as physically realistic models, it was assumed that $W'(0) = P'(0) = 0$. (See discussion, Section II-6.) Then third order polynomials were forced through the first three points of the models. For the density, these have the general form:

$$W(R) = \sum_{i=0}^{3} a_i R^i \qquad (A2)$$

and applying $W'(0) = 0$ this becomes:

$$W(R) = W(0) + a_2 R^2 + a_3 R^3 \qquad (A3)$$

This satisfies the central point identically, so a_2 and a_3 can be found by solving a second order linear system. From the results, a set of values of W can be found at interpolated values of R. A fully equivalent procedure was followed for P.

Once a_2 and a_3 have been found in (A3), M can be interpolated by integrating (A3). From (II-58) this is:

$$M = W(0)R^3 + \frac{3}{5} a_2 R^5 - \frac{1}{2} a_3 R^6 \tag{A4}$$

In general, the values of M at the second and third tabular positions do not quite agree with the results from (A4). Since M is the derived quantity here, and integration and roundoff errors must exist in the source data, it was deemed more consistent to alter the tabular values of M at these two points to agree with (A4).

All these calculations, from determining a_2 and a_3 for W and P, through interpolating the values of W, P, and M from (A3) and (A4), and finally reordering to make R monotonic, are done by the program CENDENS listed below.

```
      V CENDENS
F1]     V3+V3V2+V3V--EI2  5J
[2]     TMAT+BV2.[1.5] V3
[3]     []+A+[RM]+.×WR] 3]-W[1]
[4]     []+B+J[M]+.×P[2 3]-P[1]
[5]     W[]+W[1]+RN×RN×A[1]+AE2+×RN
[6]     R+[R,RN)[I+AP,RN]
[7]     W+[W,WN)[L]
[8]     MN+RN×RN×RN×W[1]+RN×RN×(0.6×A[1])+0.5×AE2+×RN
[9]     V3×W[1]+V2×(0.6×A[1])+0.5×A[2]×V
[10]    N+(N+WN)[L]
[11]    PN+P[1]+RN×RN×B[1]+BE2]×RN
[12]    P[]+V2+0[1]+BE2]×V
[13]    P+[P,PN)[T]
        V
```

A somewhat different approach was followed beyond the first three tabular points. If interpolation is desired at a point R in the range $R_i \leq R \leq R_{i+1}$, where the R_i are the source tabular values, then parabolic fits were made to the point sets R_{i-1}, R_i, R_{i+1} and R_i, R_{i+1}, R_{i+2}. These were evaluated at R and averaged. The function BIQUAD was written to carry this out for a given input table R_i and associated $Y(R_i)$. Provisions are included to omit one parabola if either R_{i-1} or R_{i+2} is missing. Here is BIQUAD.

```
      ∇ W←Z BIQUAD V;N;X;Y;ZV,I;J;K
[1]    X←V[⌊N←0.5×ρV]
[2]    Y←V[N+⌊N]
[3]    ←(Z⌈X)/W
[4]    ZV←1,Z,Z×Z
[5]    W←0
[6]    I←+/X÷Z
[7]    J←(⌊3)+I-2
[8]    →(K←I=1)/R
[9]    W←+/ZV×X[J] PARAB Y[J]
[10]   R:←(I ⌈N-1)/0
[11]   W←0.5×(1+K)×W+÷/ZV×X[J] PARAB Y[J←J+1]
[12]   →0
[13]   Q:W←Y[X[I]]
      ∇
```

To apply this to a set of interpolation points R_n, for all the variables W, M, and P, and finally reorder the results to again make R monotonic, BIQUAD is called by the program EXPAND, shown here:

```
      ∇ EXPAND
[1]    N←ρRN
[2]    K←1
[3]    MN←WN←PN←⍳0
[4]    Q:A←RN[K]
[5]    MN←MN,⍺ BIQUAD R,M
[6]    WN←WN,⍺ BIQUAD R,W
[7]    PN←PN,⍺ BIQUAD R,P
[8]    K←K+1
[9]    →(K≤N)/Q
[10]   R←(P,RN)[⍋P,RN]
[11]   M←(M,MN)[I]
[12]   W←(W,WN)[J]
[13]   P←(P,PN)[J]
      ∇
```

The results of all this are shown in Figure A-1 and in Tables A2-A6.
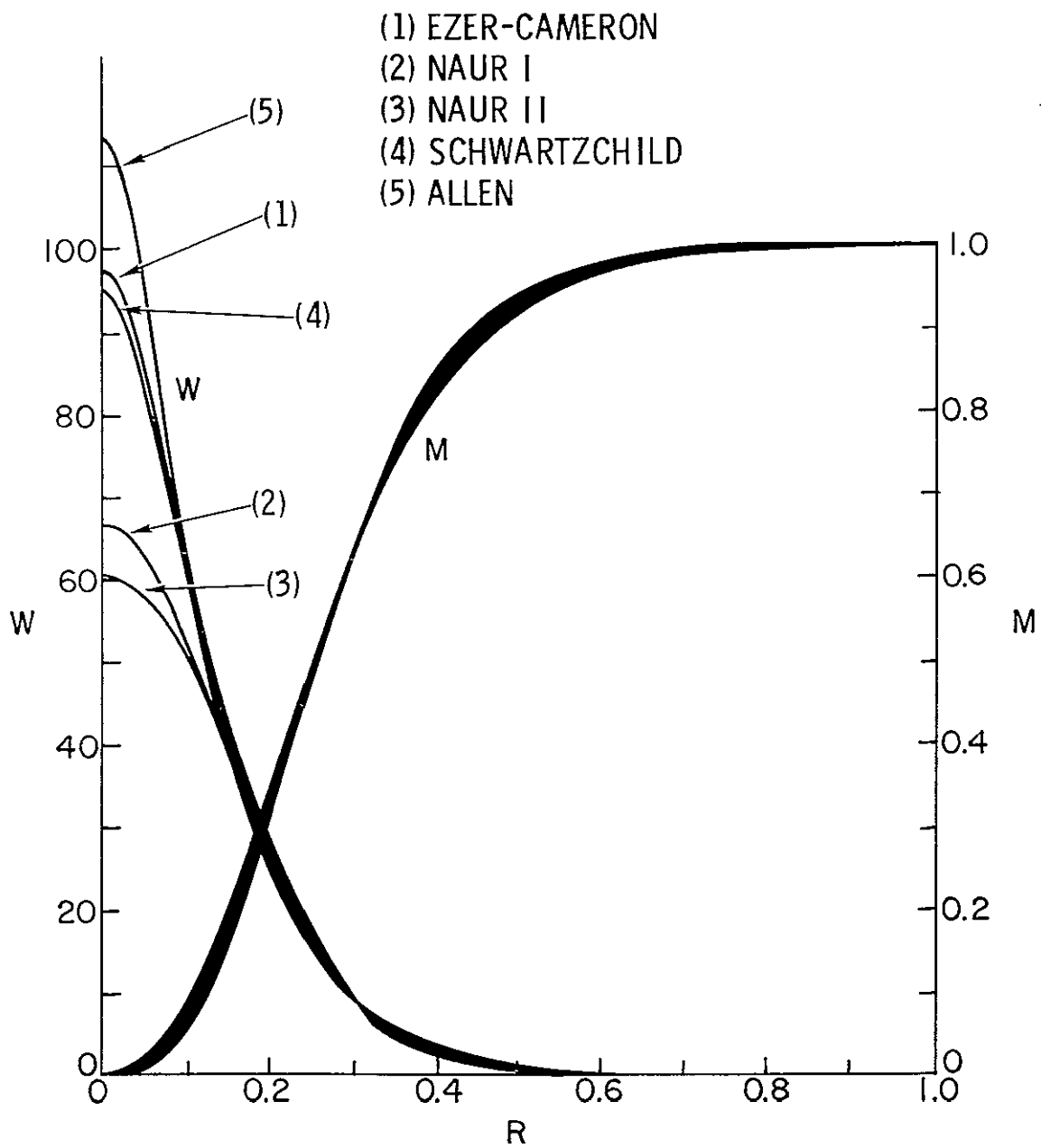
Figure A-1. Density and Mass for Published Models

3. Derived Quantities

Now that W, M, and P are available as fairly dense functions of R, we can tabulate some further necessary quantities. These include M/R, $e^{-2\Lambda}$, which we get directly from (II-34), and then $\Phi(R)$ from integrating (II-35). integrating the potential $\Phi(R)$, we know from (II-41) that:

$$\Phi(1) = -\Lambda(1) = \frac{1}{2}\ln(1 - 2\gamma) = -\gamma - \gamma^2 + O(\gamma^3) \tag{A5}$$

Thus, the integration of (II-35) can be set up as:

$$\Phi(R) = \Phi(1) - \gamma \int_R^1 e^{2\Lambda(x)}\left[\frac{M(x)}{x^2} + 3xP(x)\right]dx \tag{A6}$$

The program for doing this is POTENTIAL, which uses a simple trapezoidal rule for the integration.

```
      ∇  POIENTIAL.
[1]      K←,R
[2]      RPOT←PUT←-MUS1÷MU
[3]      PHL←1-(2:MU)×11÷R
[4]      EMU.T1:←1
[5]      PP←MUS((3×P:.P)+MF.P×R)-EMU.
[6]      Q←RP(IT←RPOT,POT×POT+0.5×(PLK-1J-RPK3^^PPTF  I1+PPCK1
[7]      K←K-1
[8]      →(K:1)/Q
[9]      PHL←MRPOT
      ∇
```

The functions M/R and $\Phi(R)$ are shown in the main Tables A2-A6. Note particularly that the earlier assumption that $\Lambda(R)$ and $\Phi(R)$ are of $O(\gamma)$ is verified for all models.

4. Cubic Coefficients

For the purpose of simplifying the calculation of gain near the central focus we have developed a first order theory, keeping only terms of $O(\gamma)$,

Table A2. Ezer-Cameron Model

R	W	M	10^6P	M/R	$-10^6\Phi$
0	97.345	0	181.23	0	11.367
.005	97.221	1.2159×10^{-5}	180.87	.00243	11.359
.01	96.852	9.7049×10^{-5}	179.81	.0097	11.351
.015	96.244	3.263×10^{-4}	178.12	.02175	11.338
.02	95.4	7.6938×10^{-4}	175.86	.03846	11.319
.025	94.325	.0014926	173.08	.0597	11.295
.03	93.025	.0025579	169.84	.08526	11.267
.035	91.505	.0040222	166.2	0.11492	11.233
.04	89.768	.0059365	162.23	0.14841	11.194
.045	87.82	.0083445	157.96	0.18543	11.151
.05	85.665	.011282	153.48	0.22564	11.104
.055	83.309	.014777	148.82	0.26867	11.052
.0611	80.166	.02	143	0.32442	10.984
.065	78.01	.023499	139.25	0.36153	10.937
.07	75.077	.028732	134.44	0.41045	10.874
.07898	69.35	.04	126.01	0.50058	10.754
.09389	62.543	.06	113.27	0.63905	10.538
0.10432	57.198	.08	102.65	0.76687	10.376
0.11475	52.808	0.1	93.45	0.87146	10.204
0.1237	49.054	0.12	85.66	0.97009	10.052
0.13264	45.682	0.14	78.58	1.0555	9.897
0.14009	42.692	0.16	72.21	1.1421	9.765
0.14754	39.956	0.18	66.62	1.22	9.63
0.15499	37.475	0.2	61.45	1.2904	9.494
0.17	32.995	0.24098	52.13	1.4176	9.219
0.19076	27.549	0.3	41.06	1.5727	8.84
0.2	25.283	0.32799	36.64	1.6399	8.673
0.22355	20.105	0.4	26.97	1.7893	8.253
0.25782	14.188	0.5	17.13	1.9393	7.667
0.29508	9.48	0.6	10.12	2.0333	7.076
0.33979	5.758	0.7	5.32	2.0601	6.439
0.3994	2.933	0.8	2.28	2.003	5.713
0.49478	0.993	0.9	0.59	1.819	4.8
1.0	0	1.0	0	1.0	2.201

Table A3. Naur I Model

R	W	M	$10^6 P$	M/R	$-10^6 \Phi$
0	66.6	0	139.01	0	10.285
.005	66.56	8.322×10^{-6}	138.93	.001664	10.278
.01	66.44	6.6503×10^{-5}	138.68	.00665	10.272
.015	66.242	2.2405×10^{-4}	138.27	.014936	10.264
.02	65.969	5.2975×10^{-4}	137.71	.026488	10.251
.025	65.623	.0010314	137	.041256	10.235
.03	65.206	.0017754	136.14	.05918	10.216
.035	64.72	.0028066	135.13	.080189	10.194
.04	64.168	.0041678	133.99	0.1042	10.168
.05	62.87	.0080408	131.3	0.16082	10.106
.06	61.332	.01369	128.09	0.22816	10.031
.07	59.569	.021363	124.41	0.30519	9.945
.08	57.598	.031259	120.28	0.39074	9.847
.091	55.211	.044877	115.25	0.49315	9.726
0.11	50.504	.076539	105.38	0.69581	9.489
0.12	47.9	.096454	99.93	0.80378	9.351
0.128	45.754	0.114	95.45	0.89063	9.235
0.135	43.883	0.13057	91.57	0.96718	9.13
0.15	39.912	0.16853	83.32	1.1236	8.897
0.165	35.993	0.20985	75.18	1.2718	8.655
0.184	31.102	0.267	65.02	1.4511	8.34
0.2	27.155	0.31909	56.78	1.5955	8.071
0.215	23.688	0.36868	49.52	1.7148	7.817
0.23	20.446	0.419	42.72	1.8217	7.564
0.276	12.787	0.561	26.68	2.0326	6.817
0.334	6.6467	0.71	13.96	2.1257	5.972
0.397	3.1235	0.826	6.46	2.0806	5.196
0.482	1.1056	0.919	2.29	1.9066	4.366
0.529	0.6334	0.948	1.25	1.7921	4
0.57	0.3743	0.968	0.83	1.6982	3.722
0.606	0.2331	0.979	0.42	1.6155	3.507

Table A3. Naur I Model (Continued)

R	W	M	$10^6 P$	M/R	$-10^6 \Phi$
0.637	0.1539	0.986	0.42	1.5479	3.339
0.664	0.1052	0.99	0.21	1.491	3.205
0.687	.0746	0.993	0.21	1.4454	3.099
0.707	.0546	0.995	0.21	1.4074	3.012
0.742	.0313	0.997	0	1.3437	2.871
0.783	.0149	0.999	0	1.2759	2.721
0.82	.0007	1.0	0	1.2195	2.598
1.0	0	1.0	0	1.0	2.123

Table A4. Naur II Model

R	W	M	$10^6 P$	M/R	$-10^6 \Phi$
0	60.8	0	131.78	0	10.176
.005	60.772	7.5979×10^{-6}	131.72	.00152	10.169
.01	60.689	6.0733×10^{-5}	131.55	.00607	10.165
.015	60.551	2.0469×10^{-4}	131.27	.01365	10.157
.02	60.36	4.8428×10^{-4}	130.87	.02421	10.145
.025	60.115	9.4355×10^{-4}	130.36	.03774	10.131
.03	59.818	.0016256	129.74	.05419	10.113
.035	59.47	.0025724	129.01	.0735	10.093
.04	59.071	.0038244	128.16	.09561	10.069
.05	58.125	.0073978	126.15	0.14796	10.012
.06	56.985	.012634	123.69	0.21056	9.943
.07	55.657	.019784	120.81	0.28263	9.863
.08	54.149	.029059	117.49	0.36324	9.772
.091	52.288	.04191	113.36	0.46055	9.66
0.11	48.539	.071473	105.81	0.64976	9.439
0.12	46.409	.090307	101.54	0.75256	9.31

R	W	M	$10^6 P$	M/R	$-10^6 \Phi$
0.128	44.627	0.107	97.97	0.83594	9.201
0.135	43.02	0.1231	94.39	0.91185	9.103
0.15	39.542	0.16039	86.6	1.0693	8.882
0.165	36.02	0.20149	78.63	1.2211	8.65
0.184	31.494	0.259	68.28	1.4076	8.347
0.2	27.738	0.31218	60.11	1.5609	8.084
0.215	24.373	0.36308	52.84	1.6888	7.835
0.23	21.158	0.415	45.95	1.8043	7.585
0.276	13.254	0.562	28.83	2.0362	6.842
0.334	6.8096	0.717	14.74	2.1467	5:993
0.397	3.1373	0.835	6.72	2.1033	5.209
0.482	1.064	0.927	2.38	1.9232	4.371
0.529	0.5861	0.956	1.3	1.8072	4.002
0.57	0.3393	0.973	0.65	1.707	3.723
0.606	0.2067	0.982	0.43	1.6205	3.506
0.637	0.1332	0.988	0.22	1.551	3.338
0.664	.0894	0.992	0.22	1.494	3.204
0.687	.062	0.994	0.22	1.4469	3.098
0.707	.0448	0.996	0	1.4088	3.011
0.742	.0251	0.998	0	1.345	2.869
0.783	.0114	0.999	0	1.2759	2.72
0.82	.0052	1.0	0	1.2195	2.597
1.0	0	1.0	0	1.0	2.122

Table A5. Schwartzchild Model

R	W	M	$10^6 P$	M/R	$-10^6 \Phi$
0	95.544	0	176.86	0	10.641
.002	95.503	7.6415×10^{-7}	176.78	.000382	10.639
.005	95.297	1.1924×10^{-5}	176.36	.002385	10.636
.008	94.935	4.8728×10^{-5}	175.63	.006091	10.632
.011	94.435	1.2626×10^{-4}	174.64	.011478	10.627
.014	93.817	2.5924×10^{-4}	173.42	.018517	10.619
.017	93.099	4.6192×10^{-4}	172.01	.027172	10.61
.02	92.301	7.4803×10^{-4}	170.47	.037401	10.599
.023	91.442	.0011308	168.83	.049165	10.586
.026	90.541	.0016229	167.14	.062419	10.572
.03	89.306	.0024701	164.88	.082335	10.55
.035	87.775	.003875	162.18	0.11071	10.519
.04	86.338	.0057131	159.82	0.14283	10.483
.044	85.048	.0074662	157.37	0.16969	10.451
.049	83.325	.010153	153.96	0.2072	10.409
.054	81.48	.01339	150.16	0.24795	10.362
.06	79.105	.018	145.09	0.3	10.301
.066	76.594	.023513	139.69	0.35625	10.235
.073	73.537	.031122	133.15	0.42633	10.151
.08	70.34	.04	126.37	0.5	10.062
.09	65.572	.055375	116.89	0.61528	9.923
0.1	60.842	.073	107.56	0.73	9.773
0.11	56.514	.092	98.68	0.83636	9.615
0.12	52.385	0.113	90.08	0.94167	9.452
0.13	48.347	0.13656	81.64	1.0505	9.283
0.14	44.484	0.162	73.56	1.1571	9.109
0.15	40.852	0.18888	65.99	1.2592	8.933
0.16	37.429	0.217	58.97	1.3562	8.754
0.17	34.207	0.24613	52.64	1.4478	8.574
0.18	31.204	0.276	46.84	1.5333	8.393
0.19	28.448	0.30631	41.49	1.6122	8.213

Table A5. Schwartzchild Model (Continued)

R	W	M	10^6P	M/R	$-10^6\Phi$
0.2	25.894	0.337	36.61	1.685	8.034
0.22	21.291	0.399	28.23	1.8136	7.681
0.24	17.426	0.46	21.61	1.9167	7.337
0.26	14.132	0.519	16.43	1.9962	7.005
0.28	11.408	0.574	12.38	2.05	6.687
0.3	9.1454	0.626	9.3	2.0867	6.384
0.32	7.298	0.672	6.93	2.1	6.098
0.34	5.8238	0.716	5.16	2.1059	5.827
0.36	4.658	0.753	3.86	2.0917	5.573
0.38	3.7085	0.788	2.88	2.0737	5.334
0.4	2.939	0.818	2.13	2.045	5.11
0.42	2.3238	0.844	1.58	2.0095	4.9
0.44	1.8501	0.867	1.18	1.9705	4.704
0.46	1.4662	0.887	0.87	1.9283	4.52
0.48	1.1647	0.904	0.65	1.8833	4.348
0.5	0.923	0.919	0.48	1.838	4.187
0.52	0.7332	0.932	0.36	1.7923	4.036
0.54	0.5797	0.943	0.27	1.7463	3.894
0.56	0.4605	0.953	0.2	1.7018	3.761
0.58	0.3649	0.961	0.15	1.6569	3.636
0.6	0.2879	0.967	0.11	1.6117	3.519
0.62	0.2281	0.973	.08	1.5694	3.408
0.64	0.1804	0.979	.06	1.5297	3.304
0.66	0.1426	0.982	.04	1.4879	3.206
0.68	0.112	0.985	.03	1.4485	3.113
0.7	.0881	0.988	.02	1.4114	3.025
0.72	.0691	0.989	.02	1.3736	2.942
0.74	.054	0.992	.01	1.3405	2.863
0.76	.042	0.994	.01	1.3079	2.788
0.78	.0326	0.995	.01	1.2756	2.717
0.8	.0252	0.996	0	1.245	2.649

Table A5. Schwartzchild Model (Continued)

R	W	M	10^6P	M/R	$-10^6\Phi$
0.82	.0195	0.997	0	1.2159	2.585
0.84	.0151	0.998	0	1.1881	2.523
0.86	.0119	0.998	0	1.1605	2.465
0.88	.0091	0.999	0	1.1352	2.409
0.9	.0067	0.999	0	1.11	2.355
0.92	.0047	1.0	0	1.087	2.304
0.94	.0029	1.0	0	1.0638	2.255
0.96	.0015	1.0	0	1.0417	2.208
0.98	.0005	1.0	0	1.0204	2.163
1.0	0	1.0	0	1.0	2.12

Table A6. Allen Third Edition Model

R	W	M	10^6P	M/R	$-10^6\Phi$
0	113.6	0	267.68	0	11.017
.002	113.56	9.086×10^{-7}	267.55	.00045	11.014
.005	113.34	1.418×10^{-5}	266.9	.00284	11.012
.01	112.59	1.1299×10^{-4}	264.66	.0113	11.003
.015	111.39	3.7885×10^{-4}	261.08	.02526	10.988
.02	109.78	8.9008×10^{-4}	256.3	.0445	10.967
.025	107.8	.0017193	250.45	.06877	10.941
.03	105.5	.0029319	243.67	.09773	10.909
.035	102.93	.0045856	236.09	0.13102	10.872
.04	100.11	.0067292	227.84	0.16823	10.83
.05	93.935	.012637	209.87	0.25274	10.731
.06	87.316	.020868	190.82	0.34779	10.616
.07	80.593	.031507	171.77	0.4501	10.487
.08	74.105	.044549	153.77	0.55687	10.345

Table A6. Allen Third Edition Model (Continued)

R	W	M	$10^6 P$	M/R	$-10^6 \Phi$
.09	68.192	.05995	137.89	0.66611	10.192
0.1	63.191	.077704	125.2	0.77704	10.031
0.13	50.69	0.14942	93.24	1.1494	9.503
0.16	40.14	0.22968	67.14	1.4355	8.936
0.2	29.11	0.35	41.46	1.75	8.185
0.25	17.577	0.49889	20.72	1.9956	7.297
0.3	9.4431	0.64	9.5	2.1333	6.497
0.35	4.8857	0.7575	3.77	2.1643	5.792
0.4	2.556	0.85	1.85	2.125	5.183
0.45	1.2314	0.9055	0.65	2.0122	4.664
0.5	0.71	0.94	0.37	1.88	4.227
0.6	0.2485	0.982	0.08	1.6367	3.539
0.7	.0561	0.994	0.01	1.42	3.035
0.8	.0128	0.999	0	1.2488	2.654
0.9	.0014	1.0	0	1.1111	2.358
1.0	0	1.0	0	1.0	2.121

and further restricting it to small R_O. This entailed the neglect of $\Lambda(R)$ and $\Phi(R)$, both of which have now been sustained for the models examined. We have also neglected the pressure, and we must now justify this. From (II-37) we can see that we can do this provided that both $P \ll W$ and $R^3 P \ll M$. This is the main reason for including P and $R^3 P$ in Tables A2-A6. Well, the tables show that $P \ll W$ everywhere, while $R^3 P \ll M$ except possibly for $R \ll 1$. In this case we know that $M(R) \sim \dot{W}(O)R^3$; so the pressure is everywhere neglectible in our models, and we conclude that the first order theory is adequate for them for small R_O.

First order theory shows that the deflection is given by (II-74), in which the parameters δ_O' and h are obtained from the model by means of (II-75, 76). Evaluation of these requires only $W(R)$ and a_2, obtained in CENDENS above. The program used is THIRD:

```
      ∇ THIRD
[1]      'ENTER MODEL NAME'
[2]      ⍺'R←R',Z←M
[3]      ⍺'W←W',Z
[4]      H←(1+R)-¯1+R
[5]      3×11 PARINT W
[6]      'ENTER A2'
[7]      Y←⎕,1+(W[1]-W)÷R*2
[8]      0.75×W[1]+H PARINT Y
      ∇
```

The integrations in THIRD are performed by a general parabolic integration PARINT, for which the theory and program are given in Appendix C. Given δ_O' and h, we can derive values for $L(O)$ and α, using (II-92) and (II-101). The results of all this are shown in Table A7. Also included there are the corresponding values for the polynomial density models, in which for realism we have taken $r_s' = r_s/2$.

Table A7. Cubic Parameters

Model	W(0)	$-a_2$	δ_0'/γ	h	$\ell(0)$ (Tm)	α
Uniform	8	0	12	6	13.67	1.5
Quadratic	20	10	20	30	8.20	4.5
Cubic	40	60	30	90	5.47	9
Ezer-Cameron	97.345	4990.6	42.642	862.98	3.5752	60.7134
Naur I	66.6	1634	38.223	422.41	4.2851	33.1536
Naur II	60.8	1121.5	36.876	349.94	4.4456	28.4689
Schwartzchild	95.544	10465	44.225	956.96	3.707	64.9153
Allen 3rd Ed.	113.6	10692	48.55	1232.2	3.3796	76.1401

Perhaps the most important result is $\ell(0)$. All these published models yield $3.37 < \ell(0) < 4.45$ Tm, which, for scale, is in the Uranus-Neptune region, though not necessarily near the ecliptic. This is not quite as restrictive as it looks since the lower limit is set by Allen's model, which is an average over four other published models. Compelling evidence may be found (or may even exist) permitting us to narrow the range of $\ell(0)$; but in its absence it would seem prudent to begin a search for $\ell(0)$ as close as 2 Tm.

An interesting contrast is provided by the polynomial models. By reducing r_s' only slightly, the cubic model would fall in the published range of W(0) and $\ell(0)$, and even the uniform density model would only need a further reduction of about 2-2.5. However, the gross mismatch in a_2, and thus h and α, would persist. We conclude that while $\ell(0)$ depends strongly only on W(0), the shape of the caustic also depends strongly on $W''(0)$.

5.   The Focal Region

A good way to compare our models is to plot the caustics together, as in Figure II-5. This time however, it is better to plot them with true dimensions, instead of normalizing to r_s and γ. To this end we can use (II-32, 89, 92, 96, 99, and 101) to rewrite the parametric description of the caustic as:

$$r = \ell(0)(1 - \alpha R_0^2)^{-1} \tag{A7}$$

$$y = - 4\gamma h r R_0^3 \tag{A8}$$

$$r - \ell(0) = \alpha r R_0^2 \tag{A9}$$

and
$$\ell(0) = \frac{r_s}{2\delta_0'} \tag{A10}$$

The arithmetic has been carried out by the program CAUSTIC:

```
    ∇ CAUSTIC
[1]    (Z; 'ENTER ' ), 'GAMMA'
[2]    G←□
[3]    Z, 'RS'
[4]    RS←□
[5]    Z, 'DOP/GAM'
[6]    L←RS÷2×D>(1-1)
[7]    Z, 'H'
[8]    A←3/(H×□)÷0
[9]    )IF<AxR2;R←L÷1-A×R2;K(X×R0^0.017×1H
[10]   Y←R×R0×R2>H×G×H
[11]   H←N;× 3  10  ;R0,(1[:?×0R×),0.001×Y
    ∇
```

using inputs from Tables A1 and A7. The results are plotted in Figure A-2. The cubic model has been added for comparison with $r_s' = r_s/2$. The curves all look much alike. Indeed, the Ezer-Cameron curve lies so close to the Allen curve that it was not plotted separately. There are strong differences in the R_0 values however; so substantial variations in the predicted gain will exist.
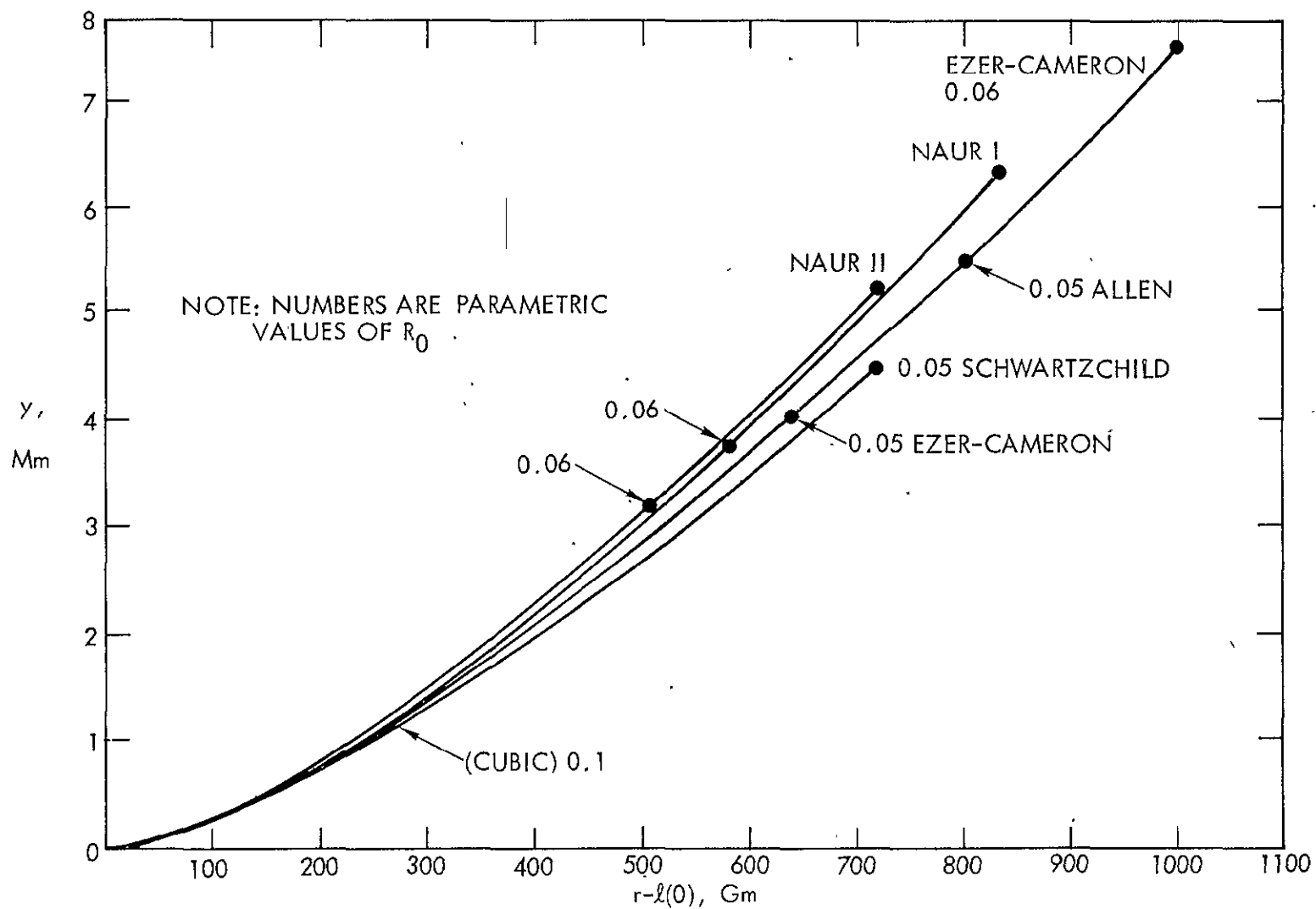
Figure A-2. Caustics for Published Solar Models

To compute the gain contours, (II-107) can be used with only minor changes:

$$G = r_s \sum \frac{R_o}{y} [1 - \frac{r}{\ell(0)} (1 - \alpha R_o^2)]^{-1}$$ (A11)

In (A11) we can specify r and y, and then solve the ray equation (II-102) for R_o, which can be restated as:

$$R_o^3 + \frac{3}{\alpha} [\frac{\ell(0)}{r} - 1]R_o - \frac{y}{2\gamma hr} = 0$$ (A12)

This gives us all we need to evaluate G.

The full procedure is an improvement over that of Section II-10, in that here we obtain the gain contours directly by selecting r and G, and then adjusting y until the solution of the third order system (A11, 12) yields the desired G. In this, (A12) is solved analytically for R_o (one or three real solutions), and then G is evaluated from (A11). Newton's method is then used to improve y until G is correct. The derivative is obtained numerically, using closely spaced y values. The program for this is GAINCON:

```
    ∇ GAINCON
[1]   (Z←'ENTER '),'DESIRED GAIN'
[2]   GD←⎕
[3]  Q:Z,'R IN TM'
[4]   K←(K1←L0÷R←⎕)-1
[5]   K2←2×K÷K1
[6]   ATH←(|A3←(A←K÷ALP)*3)*0.5
[7]   TRA←2×(|A)*0.5
[8]   B1←FGH×R
[9]   Z,'GUESSED Y IN KM'
[10]  Y←⎕
[11] Q1:→(0>D3←A3+(B←Y÷B1)*2)/Q2
[12]  R0←((B+RD3)*E)+(B-RD3←D3*0.5)*E
[13]  →Q3
[14] Q2:R0←TRA×2○PV+E×¯2○B÷ATH
[15] Q3:G←(RS÷Y)×+/|R0÷X←(K+ALP×R0*2)÷K1
[16]  Y←Y×1+(DG←G-GD)÷+/G,(X-K2)÷X*3
[17]  →(0.01<|DG)/Q1
[18]  R,G,Y
[19]  →Q
    ∇
```

This is still fairly tedious, and was actually carried out for only one model. The Ezer-Cameron model was chosen because it appears to be the most "average" of the lot. The results are shown in Figure A-3.
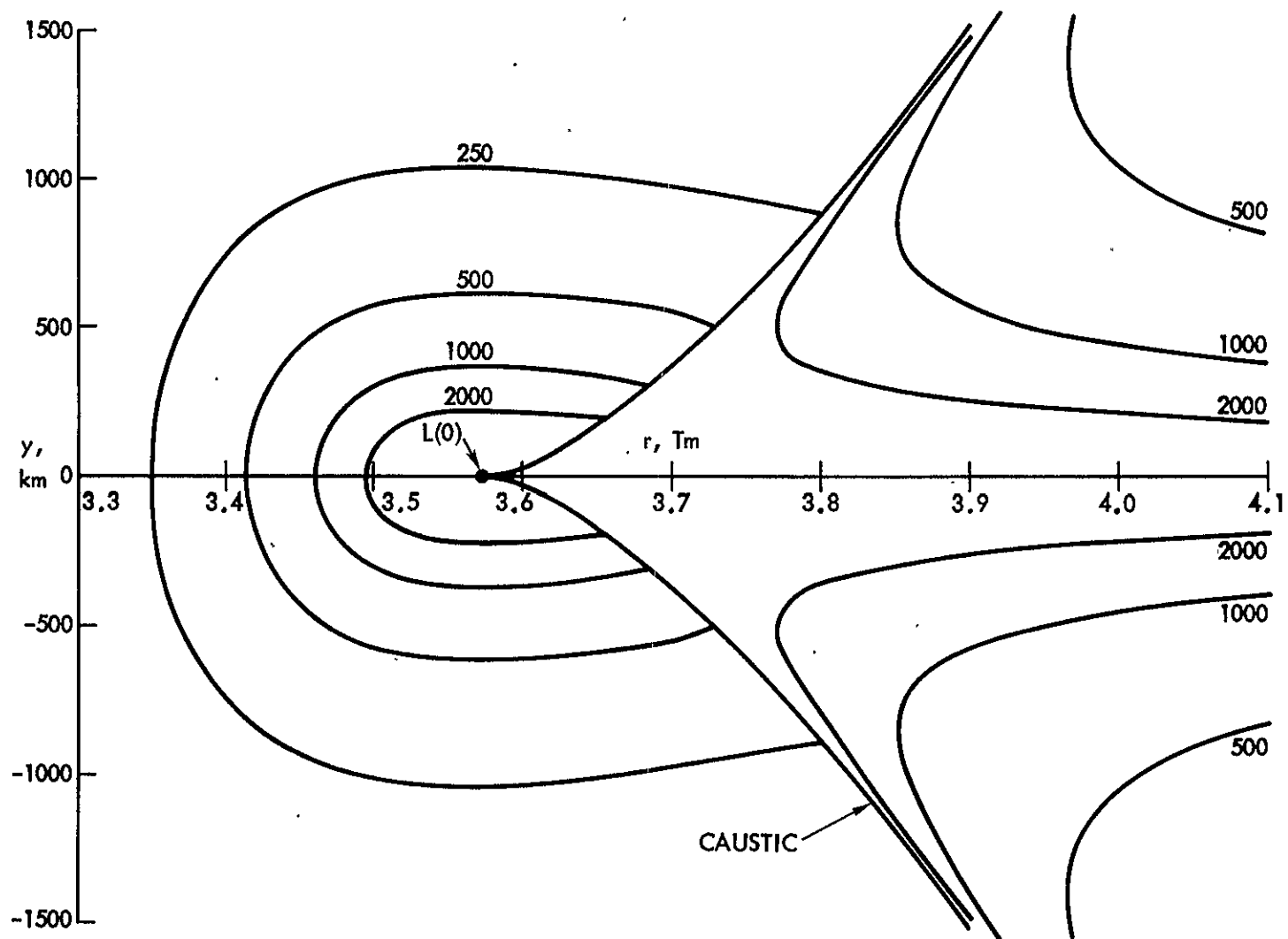
Figure A-3. Gain Contours for Ezer-Cameron Model

APPENDIX B

A SERIES SUMMATION

From the definition (II-67) we have

$$\sum_{i=1}^{\infty} c_i x^{2i} = 1 - (1 - x^2)^{\frac{1}{2}} \equiv 1 - w \tag{B1}$$

Starting from this we will establish a class of summation formulas involving the c_i. Since all $c_i > 0$, the series (B1) is uniformly convergent on the interval $0 \leq x \leq 1$, and we can integrate it term by term, on this interval. First though, we will divide through by x^{n+1}, where $n > 2$ and odd. After removing all the negative exponents from the left side this is:

$$\sum_{i=k+1}^{\infty} c_i x^{2i-n-1} = (1 - w)x^{-n-1} - \sum_{i=1}^{k} c_i x^{2i-n-1} \tag{B2}$$

where

$$k \equiv \frac{1}{2}(n - 1) \geq 1 \tag{B3}$$

The right side of (B2) now leads to a sum of $k + 2$ improper integrals; but taken together they must converge; so we can write the integration as:

$$\sum_{i=k+1}^{\infty} \frac{c_i}{2i-n} = \lim_{g \to 0} \int_g^1 \left[x^{-n-1} - \sum_{i=1}^{k} c_i x^{2i-n-1} - w x^{-n-1} \right] dx \tag{B4}$$

The last term is the most troublesome, but

$$\int_g^1 \frac{wdx}{x^{n+1}} = -\int_{x=g}^{x=1} \frac{w^2}{x^{n-1}} d\left(\frac{w}{x}\right)$$

$$= g^{-n}(1 - g^2)^{3/2} + \int_g^1 [(1 - n)x^{-n-1} + (n - 3)x^{1-n}]wdx \quad \text{(B5)}$$

and recognizing the original integral on the right:

$$n\int_g^1 \frac{wdx}{x^{n+1}} = g^{-n}(1 - g^2)^{3/2} + (n - 3)\int_g^1 \frac{wdx}{x^{n-1}} \quad \text{(B6)}$$

We could now iterate this process, since the final integral is of the same type as the first. The resulting series of functions of g will terminate when the coefficient of the right integral vanishes, which must happen for the given n values.

Now each succeeding term will be of the form $Kg^{2i-n}(1 - g^2)^{3/2}$, with $i \geq 1$, so if we expand $(1 - g^2)^{3/2}$, we will get

$$\int_g^1 \frac{wdx}{x^{n+1}} = \frac{1}{n} g^{-n} + \sum_{i=1}^{\infty} a_i g^{2i-n} \quad \text{(B7)}$$

with some set of finite coefficients a_i. This allows us to complete the integration of (B4):

$$\sum_{i=k+1}^{\infty} \frac{C_i}{n-2i} = \lim_{g \to 0} \left[\frac{1}{n}(1 - g^{-n}) + \sum_{i=1}^{k} \frac{C_i}{n-2i} (g^{2i-n} - 1) + \frac{1}{n} g^{-n} \sum_{i=1}^{\infty} a_i g^{2i-n} \right]$$

$$\text{(B8)}$$

After completing the sum on the left, this becomes:

$$\sum_{i=1}^{\infty} \frac{C_i}{n-2i} = \frac{1}{n} + \lim_{g \to 0} \left[\sum_{i=1}^{k} \frac{C_i}{n-2i} g^{2i-n} + \sum_{i=1}^{\infty} a_i g^{2i-n} \right] \quad \text{(B9)}$$

Finally, we observe that the series on the left has no infinite terms since n is odd. Further $|n - 2i| \geq 1$, so (B9) is dominated term by term by the all positive series $\sum_{i=1}^{\infty} C_i = 1$. Thus the left side of (B9) must converge, and be represented by the expression on the right. This can only happen if the negative exponent terms, when $i \leq k$, cancel identically. The terms with $i > k$ all vanish as $g \to 0$, and as there is no i such that $2i = n$, we are left with

$$\sum_{i=1}^{\infty} \frac{C_i}{n-2i} = \frac{1}{n} \tag{B10}$$

As a reminder, (B10) has been proved only for odd $n > 1$. Indeed, for $n = 1$, it is known to fail. In retrospect, it seems unlikely that (B10) is new; but it is not listed in any tables known to the author.

APPENDIX C

INTEGRAL EVALUATION

Extensive use has been made here of numerical integration (quadrature). The simplest method is the trapezoidal rule; however, for well behaved functions, improvements are possible. When the abscissae are evenly spaced, Romberg's method[24] has proved to be excellent. Unfortunately, even spacing has not been practical in much of the present work; and a different approach has been taken.

Suppose a function $y(x)$ has been determined at the points x_i; $i = 1, 2, \cdots, n+1$. Then consider the interval

$$h_i = x_{i+1} - x_i \tag{1}$$

The integral over h_i is:

$$I_i = \int_{x_i}^{x_{i+1}} y(x)dx \tag{2}$$

The method is to determine $y(x)$ by passing a parabola

$$y = a_1 + a_2 x + a_3 x^2 \tag{3}$$

through the points x_i, x_{i+1}, x_{i+2}. Then I_i is found analytically. For better accuracy, this process is repeated for x_{i-1}, x_i, x_{i+1}; and the two results are averaged. Of course, only one of these can be used for the first and last intervals.

For the first case, the coefficients obey:

$$
\begin{bmatrix} 1 & x_i & x_i^2 \\ 1 & x_{i+1} & x_{i+1}^2 \\ 1 & x_{i+2} & x_{i+2}^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} y_i \\ y_{i+1} \\ y_{i+2} \end{bmatrix} \tag{4}
$$

The determinant of this system works out to:

$$
D = (x_{i+1} - x_i)(x_{i+2} - x_i)(x_{i+2} - x_{i+1}) = h_i h_{i+1}(h_i + h_{i+1}) \tag{5}
$$

so the solution is:

$$
\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \frac{1}{D} \begin{bmatrix} x_{i+1}x_{i+2}h_{i+1} & -x_i x_{i+2}(h_i + h_{i+1}) & x_i x_{i+1}h_i \\ -(x_{i+1} + x_{i+2})h_{i+1} & (x_i + x_{i+2})(h_i + h_{i+1}) & -(x_i + x_{i+1})h_i \\ h_{i+1} & -(h_i + h_{i+1}) & h_i \end{bmatrix} \begin{bmatrix} y_i \\ y_{i+1} \\ y_{i+2} \end{bmatrix} \tag{6}
$$

Putting this into (3) and evaluating:

$$
I_i = h_i[a_1 + \frac{1}{2} a_2(x_i + x_{i+1}) + \frac{1}{3} a_3(x_i^2 + x_i x_{i+1} + x_{i+1}^2)]
$$

$$
= \frac{1}{2} h_i(y_i + y_{i+1}) + \frac{h_i^2}{6h_{i+1}} F_i \tag{7}
$$

where

$$
F_i \equiv (h_i + h_{i+1})^{-1} [h_{i+1}(y_{i+1} - y_i) - h_i(y_{i+2} - y_{i+1})] \tag{8}
$$

The first term is from the trapezoidal rule; so the second is a correction. It clearly vanishes when the three points are collinear. The same procedure can be applied to x_{i-1}, x_i, x_{i+1}. The result is:

$$I_i' = \frac{1}{2} h_i(y_i + y_{i+1}) + \frac{h_i^2}{6h_{i-1}} F_{i-1} \tag{9}$$

An APL program incorporating these ideas is:

```
      V Z←H PARTNT Y;TT,HS;HH;HL;F ,Y0;A;B
[1]    TT←+/Hx(1↓Y)+¯1↓Y
[2]    HS←(HH←1↓H)+HL←¯1↓H
[3]    F←((HH×¯1↓Y0)-HL×1↓\0←(1↓Y)-¯1↓Y)-HS
[4]    A←F×(HL*2)÷HH
[5]    B←F÷(HH×2)-HL
[6]    Z←(TT+(+/A[1]←0,¯1↓B)÷6)÷2
      V
```

Here H and Y are the h and y vectors; TT is twice the trapezoidal sum; HS is the vector $h_i + h_{i+1}$; and A and B are the correction vectors in (8) and (9) respectively. This program is exact when y is of the form (3), and in other cases it has proved to be a great improvement over the uncorrected trapezoidal rule.

In the ion drive programs there is a need to integrate a parallel set of y vectors over the same set x_i. The modified program with this extra capability is:

```
      V Z←H VPARTNT Y;TT;HS;HH;HL;Y0;A;B;B
[1]    TT←H+.×(1 0 ↓Y)+ ¯1 0 ↓Y
[2]    HS←(HH←1↓H)+HL← ¯1↓H
[3]    Y0←H(1 0 ↓Y)  ¯1 0 ↓Y
[4]    F←((0  ¯1 ↓Y0×N,HH)-(0 1 ↓Y0)×N,HL)÷(N,NCOL,NP-2)÷HS
[5]    A←F×N,(HL*2)÷HH
[6]    B←F×N,HH*2 ÷HL
[7]    Z←(TT+(1 /N ;1↓,0,(NCOL, ¯1 )↑0)÷6)÷2
      V
```

The individual y vectors are the columns of the input Y matrix. Note that the structure of Y (NP, NR) is assumed available, although this wasn't really necessary.

Another variation on PARINT is a cumulative integrator:

```
      ∇ Z←H CUMINT Y;TT,HS;HH;HL;F;YD,A;B
[1]      TT←+\H×(1↓Y)+¯1↓Y
[2]      HS←(HH←1↓H)+HL←¯1↓H
[3]      F←((HH×¯1↓YD)-HL×1↓YD←(1↓Y)-¯1↓Y)÷HS
[4]      A←F×(HL*2)÷HH
[5]      B←F×(HH*2)÷HL
[6]      Z←(TT+÷\((A.¯1↓B)+A[1],B)÷6)÷2
      ∇
```

This computes all the integrals from x_1 to each x_i; $i > 1$. It is used in finding the longitude histories of some of the optimal trajectories.

Note, finally, that in all three programs, only the intervals h_i are needed. If these are not immediately available, they can be obtained from the x_i by the APL statement

$$H←(1↓X)-¯1↓X$$

APPENDIX D

INTEGRATION STEP SIZE

Depending on the choice of computer, the ion drive optimization programs are either very time consuming or very expensive to run. In either case it is important to use the largest possible time step consistent with good accuracy. The methods used to insure this accuracy will now be discussed.

The integration technique is a fourth order Runge-Kutta, in which the time step is taken as

$$\Delta t = Kr \tag{1}$$

where r is the heliocentric radius at the last step. Normally we would expect the integration error to vary as K^4; but here we have the added complications of the integral conditions and the time varying power. The biparabolic quadrature technique of Appendix C might be expected to give errors varying as K^3, while the power sampling error should go as K^2. The latter is small, partly because the power decay is slow, and partly because it is sampled twice in each Δt. Overall, we might expect the errors to vary as K^n for some n.

For very small K we might expect round off errors to become important. For a very crude estimate of these, consider a run of ION2D at K = 0.1 and t_a = 5 years. This requires 84 steps, which takes about 22 minutes on an IBM 5100 for each base run. The number of arithmetic operations is thus around 10^6. Since APL generally carries about 16 floating decimals, a random walk would lead to an error of about one part in 10^{13}, and even a fully one sided error would be no worse than one part in 10^{10}. Thus, for practical values of K, round off can be ignored.

Consider some output parameter y of the program. If we iterate to completion for each of several values of K, we would expect the results to be approximated by

$$y_T = a + bK^n \qquad (2)$$

for some values of a, b, and n. If 3 or more values of K have been run, we have an easy problem in parameter identification. If this has been done, the maximum value of K is that which gives the largest tolerable error $y_T - a$.

Suppose the data consists of the sets y_i vs. K_{Ti}. Then a least squares performance index for goodness of fit is:

$$J = \frac{1}{2} \sum_i (y_{Ti} - y_i)^2 \qquad (3)$$

At best fit we require

$$\partial_\lambda J = \sum_i (y_{Ti} - y_i)\partial_\lambda y_{Ti} = 0 \qquad (4)$$

for $\lambda = a$, b, or n. Substituting in (2) these become:

$$\begin{bmatrix} 4 & \sum_i K_i^n & \sum_i y_i \\ \sum_i K_i^n & \sum_i K_i^{2n} & \sum_i y_i K_i^n \\ \sum_i K_i^n \ln K_{Ti} & \sum_i K_i^{2n} \ln K_i & \sum_i y_i K_i^n \ln K_i \end{bmatrix} \begin{bmatrix} a \\ b \\ -1 \end{bmatrix} = 0 \qquad (5)$$

Suppose we regard these relations as 3 straight lines in a, b space. Then a solution consists of a value of n for which the lines are concurrent, for which a necessary condition is that the determinant of the matrix in (5) vanishes. Thus the problem reduces to the solution of a single transcendental relation in n, after which a and b can be found by simple linear methods.

Newton's method has been chosen to find n, in which the derivative is obtained numerically. This requires that the determinant be evaluated twice, for closely spaced values of n, in each iteration. Because APL encourages matrix methods, the determinant is evaluated by means of the permutation tensor:

$$|A| = \varepsilon_{\alpha\beta\gamma} A_{1\alpha} A_{2\beta} A_{3\gamma} \qquad (6)$$

All this is carried out by the programs PFIT and FD:

```
      ∇ PFIT
[1]      P← 3 3 3 ρ0
[2]      P[1;2;3]←P[2;3;1]←P[3,1;2]←J←¯1
[3]      P[1,3,2;1←P[2,1;,3]←P[3;2;1]← 1
[4]      A← 3 3 ρ0
[5]      A[1; ]← 3]←4,+/Y
[6]      LX← ⍟X
[7]      N←0.Γ0 N
[8]      D1←FD N+1
[9]      N←N+0N←0×F←D−D1
[10]     →((1E¯6 :1DN)∧29 J←J+1)/0
[11]     J,0,N,V←⍵1 2 ,⍟⍵ 2 2 ↑A
[12]     Y←V[1]←VE2]×N
      ∇
```

```
      ∇ U←FD N
[1]      A[1;2]←A[2;1]←+/XN←Y×N
[2]      A[2;2]←+/X2N←XN×2
[3]      A[2;3]←+/YXN←Y×XN
[4]      A[3;3]←(3 +ρ×N,X2N,YXN)+, LX
[5]      U←((P+,×A[3;])+,×A[2;])+,×A[1;:]
      ∇
```

To see how well this all works, a sequence was run in ION2D with $t_a = 1.8$ years, with these results (Table D1):

Table D1. A 2D Time Step Sequence

K - Years/Tm	0.3	0.2	0.15	0.1
ω_1 - Gm/yr^2 - wt	2.3436	2.3357	2.3337	2.3329
v_{Hr} - Gm/yr	37.927	37.702	37.656	37.64
$v_{H\phi}$ - Gm/yr	354.69	354.25	354.14	354.1
P_D - watts	3193.9	3203.5	3205.9	3206.8
Δv_D - Gm/yr	252.27	251.94	251.86	251.83
m_D - kg	4444.5	4456.1	4458.9	4460
m_a - kg	1081.7	1085.3	1086.3	1086.6
J - kg	235.63	236.73	237.01	237.11
$\Delta\phi$ - rad	1.9915	1.9791	1.9741	1.9701

On putting these numbers into PFIT we find:

Table D2. PFIT Output

	a	b	n	Largest Residual
ω_1	2.3325	0.46244	3.0962	0.00010
v_{Hr}	37.634	22.917	3.6196	0.0019
$v_{H\phi}$	354.08	28.151	3.1799	0.0075
P_D	3207.3	-602.17	3.1616	0.096
Δv_D	251.81	22.079	3.2238	0.0068
m_D	4460.5	-759.39	3.204	0.14
m_a	1086.8	-208.04	3.0761	0.11
J	237.17	-68.569	3.1543	0.013
$\Delta\phi$	1.9658	0.18206	1.6254	0.000037

There are several messages here. First, from the residuals, the model (2) is an excellent fit to the data. Second, except for $\Delta\phi$, $n > 3$, indicating that in spite of the quadrature and power sampling errors, the theoretical performance of the Runge-Kutta is nearly reached. This is also a very good test of the ION2D structure. Finally, we can see that if an accuracy of, say, 1% is desired, we should choose $K \sim 0.3$.

A second test was later run in ION3D for the case t_a = 7 years, θ_a = 80 deg using the correct (1), with these results:

Table D3.   A 3D Time Step Sequence

K - Years/Tm	0.3	0.2	0.15	0.1
ω_1 - Gm/yr^2 - wt	0.20025	0.20026	0.20026	0.20026
$v_{H\rho}$ - Gm/yr	-8.5134	-8.5109	-8.5104	-8.5102
$v_{H\phi}$ - Gm/yr	47.013	47.014	47.014	47.014
v_{Hz} - Gm/yr	-11.106	-11.105	-11.105	-11.105
P_D - watts	10386	10385	10385	10385
ω_4 - Mm/yr^2 - wt	30.251	30.255	30.257	30.257
Δv_D - Gm/yr	105.36	105.36	105.36	105.36
m_D - kg	13499	13499	13499	13499
m_a - kg	9717.6	9717.7	9717.7	9717.7
J - kg	6966.5	6966.5	6966.6	6966.5
$\Delta\phi$ - rad	5.2265	5.2265	5.2268	5.2266

It is evident that for practical purposes, K = 0.3 is still adequate. Actually, the extra printout precision in $v_{H\rho}$ permits a test of the theory, and a PFIT run on this yielded a = -8.5101 Gm/yr, b = -0.24424 Gm/yr, n = 3.5832, and a largest residual of 5.67 km/yr - a perfect fit.

REFERENCES

1. Allen, C. W., "Astrophysical Quantities", 3rd Ed., Athlone Press, 1973.

2. Anderson, A. J., "Probability of Long Period (VLF) Gravitational Radiation", Nature, Vol. 229, 1971, P. 547.

3. Anderson, A. J., "Detection of Gravitational Waves by Spacecraft Doppler Data", Intl. Symp. on Experimental Gravitation, Pavia, Italy, 9-76.

4. Anderson, J. D., "Lectures on Physical and Technical Problems Posed by Precision Radio Tracking", in (11) below.

5. Arp, H., "Anomalous Redshifts in Galaxies and Quasars", Proc. Intl. Conf. "The Evolution of the Galaxies and their Cosmological Implications", Paris, France, 9-76.

6. Bahcall, J. N., and Ostriker, J. P., "Massive Black Holes in Globular Clusters?", Nature, Vol. 256, 7-3-75, P. 23.

7. Bahcall, J. N., and Ostriker, J. P., "X-Ray Pulses from Globular Clusters", Nature, Vol. 262, 7-1-76, P. 37.

8. Bahcall, N. A., Lasker, B. M., and Wamsteker, W., "The Central Region of the X-Ray Globular Cluster NGC 1851", Ap. J., 213:L105, 5-1-77.

9. Baker, R. M. L., "Astrodynamics; Applications and Advanced Topics", Academic Press, 1967.

10. "Basic Facts About Titanium", Reactive Metals Inc. brochure 1000-6502.

11. Bertotti, B. (Ed), "Experimental Gravitation", Proc of the Intl. School of Physics "Enrico Fermi", Varenna on Lake Como, 17-29 July 1972, Academic Press, 1974.

12. "Beryllium Properties and Products", The Beryllium Corp. Bulletin 2100, 1965.

13. Bolz, R. E., and Tuve, G. L., "Handbook of Tables for Applied Engineering Science", The Chemical Rubber Co. 1970.

14. Braginskii, V. B., "Gravitational Radiation and the Prospect of its Experimental Discovery", Soviet Physics Uspekhi, Vol. 8, No. 4, Jan. - Feb. 1966.

15. Braginskii, V. B., "The Detection of Gravitational Waves and Quantum Nondisturbtive Measurements", in (26) below.

16. Braginskii, V. B., et al., "An Upper Limit on the Density of Gravitational Radiation of Extraterrestrial Origin", Sov. Phys. - JETP, Vol. 39, No. 3, 9-74.

17. Bryson, A. E., and Ho, Y.-C., "Applied Optimal Control", Halstead Press (Wiley), Revised Printing, 1975.

18. Carrol, W., and Landel, R., "Solar Sail Materials", Presented at (81) below.

19. Centaur Technical Handbook", General Dynamics Rpt. GD/C-BMP64-01-1, Rev. B, 1-24-66.

20. Chinberg, V. L., and Perkins, H., "Preliminary Studies of LOX Expulsion Bladders", NASA TM X-53005, N65-12314, 10-12-64.

21. Cuffe, J. P. B., "RL-10 Derivatives for IUS/TUG", AAS 75-155, A76-12907.

22. Cuffe, J. P. B., Private Conversation, 1-3-77.

23. Cyranski, J. F., and Lubkin, E., "A Theoretical Experiment Using Gravitational Optics to Find the Density Function of a Star", An. Phys. 87, 1974, P. 205.

24. Dahlquist, G., and Björck, Å., "Numerical Methods", Prentice-Hall, 1974.

25. Davies, R. W., "Issues in Gravitational Wave Detection with Space Missions", Transactions of the Intl. Conf. on Gravitational Waves and Radiations, Paris, 1973, P. 33.

26. De Sabbata, V., and Weber, J. (Eds), "Topics in Theoretical and Experimental Gravitation Physics", Proceedings of the Intl. School of Cosmology and Gravitation, held in Erice, Trapani, Sicily, 13-25 March 1975, Plenum Press, 1977.

27. Dixon, W. J., "Sail Deployment and Control", Presented at (81) below.

28. Douglass, D. H., "Sources of Continuous Gravitational Radiation and Detection by Doppler Tracking of Spacecraft", University of Rochester, Unpublished.

29. Douglass, D. H. et al., "Two-Detector-Coincidence Search for Bursts of Gravitational Radiation", Phys. Rev. Lett., Vol. 35, No. 8, 8-25-75, P. 480.

30. Drever, R. W. P., "Gravitational Wave Astronomy", Quarterly J. Roy. Astron. Soc. (1977) 18, P. 9.

31. Drever, R. W. P. et al., "Search for Short Bursts of Gravitational Radiation", Nature, Vol. 246, 12-7-73, P. 340.

32. Drever, R. W. P. et al., "On Gravitational Radiation Detectors using Optical Sensing Techniques", Intl. Symp. on Experimental Gravitation, Pavia, Italy, 9-76.

33. Drexler, K. E., "Metal Film Sail Materials", Presented at (81) below.

34. Dwight, H. B., "Tables of Integrals and Other Mathematical Data", 4th Ed., Macmillan, 1961.

35. "Effects of Stretch and Heat Set Temperature and Time on Cryogenic Mechanical Properties of Poly/Ethyl Terephthalate Film", NASA-TN-D-4762, N68-33034.

36. Ehricke, K. A., "Space Flight, II, Dynamics", Van Nostrand, 1962. Sec. 7-7, P. 668 et seq.

37. Einstein, A., "The Foundation of the General Theory of Relativity", (Translation); Annalen der Physik, 49, 1916.

38. Estabrook, F. B., and Wahlquist, H. D., "Response of Doppler Spacecraft Tracking to Gravitational Radiation", General Relativity and Gravitation, Vol. 6, No. 5, 1975, P. 439.

39. Ezer, D., and Cameron, A. G. W., "A Study of Solar Evolution", Can. J. Phys. 43, 8, 1965, P. 1497.

40. Fairbank, W. M., "The Use of Low-Temperature Technology in Gravitational Experiments", in (11) above.

41. Fairbank, W. M., et al., "Search for Gravitational Radiation Using Low-Temperature Techniques", in (11) above.

42. General Physical Constants Recommended by NAS-NRC, NBS Misc. Publ. 253, Rev. 5-69, U.S. Govt. P.O.

43. Hamilton, W. O., et al., "The L.S.U. Low Temperature Gravity Wave Experiment", in (26) above.

44. Harris, L. A., Suer, H. S., Skene, W. T., and Benjamin, R. J., "The Stability of Thin-Walled Unstiffened Circular Cylinders Under Axial Compression Including the Effects of Internal Pressure", J. of the Aeronautical Sciences, Vol. 24, 8-57, P. 587.

45. Hodgeman, C. D., et al., "Handbook of Chemistry and Physics", 39th Ed., Chemical Rubber Publ. Co., 1957-8.

46. Hoggatt, J. T., "Investigation of the Feasibility of Developing Low Permeabilit Polymeric Films", NAS CR 120892, N72-25159, 12-71.

47. Hollenbeck, G. R., "A Method of Deploying Solar Sails", Presented at (81) below.

48. Hough, J., et al., "Search for Continuous Gravitational Radiation", Nature, Vol. 254, 4-10-75, P. 498.

49. Hudson, T., The Marquardt Co., Private communication, 1-24-77.

50. Ipser, J. R., "Gravitational Radiation from Slowly Rotating, Fully Relativistic Stars", Ap. J. 166, 5-15-71, P. 175.

51. Jaki, S. L., "A Forgotten Bicentenary: Johann Georg Soldner", Sky and Telescope, Vol. 55, No. 6, 6-78.

52. Jane's, "All the World's Aircraft", 1975-6.

53. Jones, J. B., Caltech J.P.L., Private communication, 2-4-77.

54. Kafka, P., "Optimum Detection of Signals through Linear Devices with Thermal Noise Sources and Application to the Munich-Frascati Weber-type Gravitational Wave Detectors", in (26) above.

55. Kaufman, J. G., and Wanderer, E. T., "Aluminum for Cryogenic Applications", Machine Design, 11-11-65, P. 199.

56. Lawden, D. F., "Optimal Programming of Rocket Thrust Direction", Astronautica Acta, Vol. I, 1955, P. 41.

57. Lawden, D. F., "Optimal Escape from a Circular Orbit", Astronautica Acta, Vol. IV, 1958, P. 218.

58. Lee, M., et al., "Gravitational-radiation-detector observations in 1973 and 1974", Phys. Rev. D. Vol. 14, No. 4, 8-15-76, P. 893.

59. Machine Design, Plastics-Elastomers Reference Issue, 2-15-73.

60. MacNeal, R. H., "Heliogyro Stability and Control", Presented at (81) below.

61. "Materials Selector 77", Materials Engineering, Vol. 8, No. 6, Mid-November 1976.

62. Melbourne, W. G., and Curkendall, D. W., "Radio Metric Direction Finding: A New Approach to Deep Space Navigation", AAS/AIAA Astrodynamics Specialist Conference, Jackson Hole, Wyoming, 9-7-77.

63. Melbourne, W. G., and Sauer, C. G. Jr., "Optimum Thrust Programs for Power Limited Propulsion Systems", Caltech J.P.L. TR 32-118, 6-15-61.

64. Melbourne, W. G., and Sauer, C. G. Jr., "Optimum Interplanetary Rendezvous Trajectories with Power Limited Vehicles", Caltech J.P.L. TR 32-226, Rev. 1, 10-15-62.

65. Mironovskii, V. N., "Gravitational Radiation of Double Stars", Soviet Astronomy, Vol. 9, No. 5, March-April 1966.

66. Misner, C. W., Thorne, K. S., and Wheeler, J. A., "Gravitation", Freeman 1973.

67. Moss, G. E., Miller, L. R., and Forward, R. L., "Photon-Noise-Limited Laser Transducer for Gravitational Antenna", Applied Optics, Vol. 10, No. 11, 11-71, P. 2495.

68. Naur, P., "Two Models for the Interior of the Sun", Ap. J. 119, 1954, P. 365.

69. Paik, H., "Low-Temperature Gravitational Wave Detector", in (11) above.

70. Peters, P. C., and Mathews, J., "Gravitational Radiation from Point Masses in a Keplerian Orbit", Phys. Rev., Vol. 131, No. 1, 7-1-63.

71. Pizzella, G., "Estimated Sensitivity of the Low Temperature Gravitational Wave Antenna in Rome", in (26) above.

72. Press, W. H., and Thorne, K. S., "Gravitational Wave Astronomy", Annual Review of Astronomy and Astrophysics, Vol. 10, 1972, P. 355.

73. "Reference Data for Radio Engineers", 4th Ed., ITT Corp., 1956.

74. Rosen, N., "Bimetric Theory of Gravitation", in (26) above.

75. Rouklove, P., Caltech J.P.L., Private communication, 12-2-77.

76. Sargent, W. L. W., et al., "Dynamical Evidence for a Central Mass Concentration in the Galaxy M87", Ap. J. 221, No. 3, Part 1, 5-1-78, P. 731.

77. Sauer, C. G. Jr., "Optimum Solar-Sail Interplanetary Trajectories", AIAA/AAS Astrodynamics Conf., San Diego, Cal., 8-18-76, Paper No. 76-792.

78. Schwartzchild, M., "Structure and Evolution of the Stars", Princeton Univ. Press, 1958, Dover, 1965.

79. Scott, R. B., "Cryogenic Engineering", Van Nostrand, 1959.

80. Seifert, H. S., Private discussions, fall 1976.

81. Solar Sailing Small Symposium, at Caltech J.P.L., 4-20-77, C. B. Uphoff and P. Penzo, J.P.L., Chairmen.

82. Sonnabend, D., "Drag-Free Air Density Instrument Requiring No Propellant", Stanford Univ. Guidance and Control Lab., 4-76.

83. Sonnabend, D., "Spectral Density of Long-Baseline Gravitational Wave Detectors", Caltech J.P.L. IOM 312/78.7-36, 9-29-78.

84. "Space Shuttle", NASA SP-407, 1976.

85. Stratton, H., Rocketdyne Div. of Rockwell Intl., Private communication, 1-27-77.

86. Sullivan, F. R., Faulkner, G. E., and Clauss, F. J., "Glue-Spot Welding Joins Aluminum Stronger than Spot, Cheaper than Fusion", Materials Engineering, 1-71, P. 14.

87. Thorne, K. S., "The Generation of Gravitational Waves: A Review of Computational Techniques", in (26) above.

88. Thorne, K. S., and Braginskii, V. B., "Gravitational-Wave Bursts from the Nuclei of Distant Galaxies and Quasars: Proposal for Detection using Doppler Tracking of Interplanetary Spacecraft", Ap. J. 204:L1, 2-15-76.

89. Vaccari, J. A., "New Findings Stretch Titanium Performance Limits", Materials Engineering, 5-67, P. 94.

90. Wagoner, R. V., "Test for the Existence of Gravitational Radiation", Ap. J. 196, L63, 3-1-75.

91. Wagoner, R. V., and Will, C. M., "Post-Newtonian Gravitational Radiation from Orbiting Point Masses", Ap. J. 210, 12-15-76, P. 764.

92. Weber, J., "Evidence for Discovery of Gravitational Radiation", Phys. Rev. Lett., Vol. 22, No. 24, 6-16-69, P. 1320.

93. Weber, J., "Gravitational Radiation Experiments", Phys. Rev. Lett., Vol. 24, No. 6, 2-9-70, P. 276.

94. Weinberg, S., "Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity", Wiley, 1972.

95. Will, C. M., "Gravitational Radiation from Binary Systems in Alternative Metric Theories of Gravity: Dipole Radiation and the Binary Pulsar", Stanford Univ. Inst. of Theoretical Physics #543, 9-76.

96. Wood, K. D., "Aerospace Vehicle Design, Vol. II", Johnson Publ. Co., 1964.

97. Wright, J. L., and Warmke, J. M., "Solar Sail Mission Applications", AIAA/AAS Astrodynamics Conf., San Diego, Cal., Paper 76-808, 8-18-76.

98. Yariv, A., "Introduction to Optical Electronics", 2nd Ed., Holt, Rinehart and Winston, 1976.

99. Young, P. J., et al., "Evidence for a Supermassive Object in the Nucleus of the Galaxy M87 from SIT and CCD Area Photometry", Ap. J. 221, No. 3, Part 1, 5-1-78, P. 721.

100. Zimmerman, M., "Gravitational Radiation from the Crab and Vela Pulsars - A Revised Estimate", Orange Aid Preprint 508, 11-77.